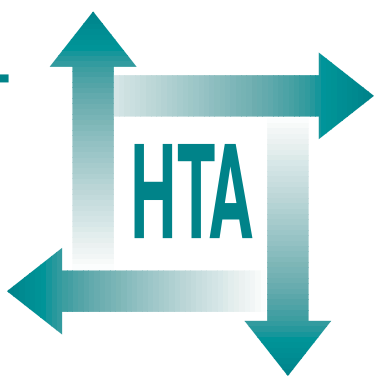
Health Technology Assessment 1998; Vol. 2: No. 9

Review

Screening for speech and language delay: a systematic review of the literature

J Law
J Boyle
F Harris
A Harkness
C Nye

Health Technology Assessment
NHS R&D HTA Programme



Standing Group on Health Technology

Chair: Professor Sir Miles Irving,
Professor of Surgery, University of Manchester, Hope Hospital, Salford †

Dr Sheila Adam,
Department of Health

Professor Martin Buxton,
Professor of Economics, Brunel University †

Professor Angela Coulter,
Director, King's Fund, London

Professor Anthony Culyer,
Deputy Vice-Chancellor, University of York

Dr Peter Doyle,
Executive Director, Zeneca Ltd,
ACOST Committee on Medical Research & Health

Professor John Farndon,
Professor of Surgery, University of Bristol †

Professor Charles Florey,
Department of Epidemiology & Public Health, Ninewells Hospital & Medical School, University of Dundee †

Professor John Gabbay,
Director, Wessex Institute for Health Research & Development †

Dr Tony Hope,
The Medical School, University of Oxford †

Professor Howard Glennester,
Professor of Social Science & Administration, London School of Economics & Political Science

Professor Sir John Grimley Evans,
Department of Geriatric Medicine, Radcliffe Infirmary, Oxford †

Mr John H James,
Chief Executive, Kensington, Chelsea & Westminster Health Authority

Professor Richard Lilford,
Regional Director, R&D, West Midlands †

Professor Michael Maisey,
Professor of Radiological Sciences, UMDS, London

Dr Jeremy Metters,
Deputy Chief Medical Officer, Department of Health †

Mrs Gloria Oates,
Chief Executive, Oldham NHS Trust

Dr George Poste,
Chief Science & Technology Officer, SmithKline Beecham †

Professor Michael Rawlins,
Wolfson Unit of Clinical Pharmacology, University of Newcastle-upon-Tyne

Professor Martin Roland,
Professor of General Practice, University of Manchester

Mr Hugh Ross,
Chief Executive, The United Bristol Healthcare NHS Trust †

Professor Ian Russell,
Department of Health, Sciences & Clinical Evaluation, University of York

Professor Trevor Sheldon,
Director, NHS Centre for Reviews & Dissemination, University of York †

Professor Mike Smith,
Director, The Research School of Medicine, University of Leeds †

Dr Charles Swan,
Consultant Gastroenterologist, North Staffordshire Royal Infirmary

Dr John Tripp,
Department of Child Health, Royal Devon & Exeter Healthcare NHS Trust †

Professor Tom Walley,
Department of Pharmacological Therapeutics, University of Liverpool †

Dr Julie Woodin,
Chief Executive, Nottingham Health Authority †

† Current members

HTA Commissioning Board

Chair: Professor Charles Florey, Department of Epidemiology & Public Health,
Ninewells Hospital & Medical School, University of Dundee †

Professor Ian Russell,
Department of Health, Sciences & Clinical Evaluation, University of York *

Dr Doug Altman,
Director, Institute of Health Sciences, Oxford †

Mr Peter Bower,
Independent Management Consultant, Newcastle-upon-Tyne †

Ms Christine Clark,
Hon. Research Pharmacist, Hope Hospital, Salford †

Professor David Cohen,
Professor of Health Economics, University of Glamorgan

Mr Barrie Dowdeswell,
Chief Executive, Royal Victoria Infirmary, Newcastle-upon-Tyne

Professor Martin Eccles,
Professor of Clinical Effectiveness, University of Newcastle-upon-Tyne †

Dr Mike Gill,
Brent & Harrow Health Authority †

Dr Jenny Hewison,
Senior Lecturer, Department of Psychology, University of Leeds †

Dr Michael Horlington,
Head of Corporate Licensing, Smith & Nephew Group Research Centre

Professor Sir Miles Irving
(Programme Director), Professor of Surgery, University of Manchester, Hope Hospital, Salford †

Professor Alison Kitson,
Director, Royal College of Nursing Institute †

Professor Martin Knapp,
Director, Personal Social Services Research Unit, London School of Economics & Political Science

Dr Donna Lamping,
London School of Hygiene & Tropical Medicine †

Professor Theresa Marteau,
Director, Psychology & Genetics Research Group, UMDS, London

Professor Alan Maynard,
Professor of Economics, University of York †

Professor Sally McIntyre,
MRC Medical Sociology Unit, Glasgow

Professor Jon Nicholl,
Director, Medical Care Research Unit, University of Sheffield †

Professor Gillian Parker,
Nuffield Professor of Community Care, University of Leicester †

Dr Tim Peters,
Department of Social Medicine, University of Bristol †

Professor David Sackett,
Centre for Evidence Based Medicine, Oxford

Professor Martin Severs,
Professor in Elderly Health Care, Portsmouth University †

Dr David Spiegelhalter,
MRC Biostatistics Unit, Institute of Public Health, Cambridge

Dr Ala Szczepura,
Director, Centre for Health Services Studies, University of Warwick †

Professor Graham Watt,
Department of General Practice, Woodside Health Centre, Glasgow †

Professor David Williams,
Department of Clinical Engineering, University of Liverpool

Dr Mark Williams,
Public Health Physician, Bristol

Dr Jeremy Wyatt,
Institute for Health Sciences, University College London †

* Previous Chair
† Current members

How to obtain copies of this and other HTA Programme reports.
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

Screening for speech and language delay: a systematic review of the literature

J Law[1]
J Boyle[2]
F Harris[1]
A Harkness[2]
C Nye[3]

[1] City University, London, UK
[2] Strathclyde University, Glasgow, UK
[3] University of Central Florida, Florida, USA

Published July 1998

This report should be referenced as follows:

Law J, Boyle J, Harris F, Harkness A, Nye C. Screening for speech and language delay: a systematic review of the literature. *Health Technol Assessment* 1998;**2**(9).

Health Technology Assessment is indexed in *Index Medicus*/MEDLINE and *Excerpta Medica*/EMBASE. Copies of the Executive Summaries are available from the NCCHTA web site (see overleaf).

NHS R&D HTA Programme

The overall aim of the NHS R&D Health Technology Assessment (HTA) programme is to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and work in the NHS. Research is undertaken in those areas where the evidence will lead to the greatest benefits to patients, either through improved patient outcomes or the most efficient use of NHS resources.

The Standing Group on Health Technology advises on national priorities for health technology assessment. Six advisory panels assist the Standing Group in identifying and prioritising projects. These priorities are then considered by the HTA Commissioning Board supported by the National Coordinating Centre for HTA (NCCHTA).

This report is one of a series covering acute care, diagnostics and imaging, methodology, pharmaceuticals, population screening, and primary and community care. It was identified as a priority by the Population Screening Panel.

The views expressed in this publication are those of the authors and not necessarily those of the Standing Group, the Commissioning Board, the Panel members or the Department of Health. The editors wish to emphasise that funding and publication of this research by the NHS should not be taken as implicit support for the recommendations for policy contained herein. In particular, policy options in the area of screening will, in England, be considered by the National Screening Committee. This Committee, chaired by the Chief Medical Officer, will take into account the views expressed here, further available evidence and other relevant considerations.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

Series Editors: Andrew Stevens, Ruairidh Milne and Ken Stein
Assistant Editor: Jane Robertson and Jane Royle

The editors have tried to ensure the accuracy of this report but cannot accept responsibility for any errors or omissions. They would like to thank the referees for their constructive comments on the draft document.

ISSN 1366-5278

Published by Core Research, Alton, on behalf of the NCCHTA.
Printed on acid-free paper in the UK by The Basingstoke Press, Basingstoke.

Copies of this report can be obtained from:

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 1703 595 639 Email: hta@soton.ac.uk
http://www.soton.ac.uk/~hta

Contents

Glossary and list of abbreviations

Technical terms and abbreviations are used throughout this report. The meaning is usually clear from the context but a glossary is provided for the non-specialist reader. In some cases usage differs in the literature but the term has a constant meaning throughout this review.

Glossary

ADD Attention Deficit Disorder (also known as Attention Deficit Hyperactivity Disorder). Very poor listening and attention skills, often associated with speech and language impairment.

Aetiology Causes of illness or disorder.

Analysis of variance A statistical technique for comparing differences between groups.

Articulation The physical movements of the mouth and throat involved in making the different speech sounds.

Coder A person who reads and extracts information from research reports.

Confidence interval The interval within which the population mean is expected to lie.

Construct validity The ability of a measure or an intervention to reflect an underlying construct. In clinical terms this may be the capacity to assess accurately the logical consequences of a disorder.

Criterion-referenced test An assessment measuring performance against a set criterion. (For example, performance score may be expressed as skills achieved.) Compare norm-referenced measure.

Cut-off Term used for a critical score on an assessment, marking the boundary between those scores considered as 'pass' and those considered as 'fail'.

Didactic intervention approaches Where the child is given a model by the adult, who makes a direct attempt to elicit the production of the modelled item by the child.

Dyslexia A language disorder reducing the ability of reading and writing. Sometimes referred to as a specific learning difficulty.

Dysfluency Condition where speech is produced with hesitations or repetitions, such as interrupt the usual flow of speech. Some dysfluency is normal within pre-school speech and language development. See also stammering/stuttering.

Effect modifiers Factors such as age or disease severity, which alter the effectiveness of an intervention.

Effect size Generic term for statistics used to measure the change in performance of a treated group over and above that of an untreated group, when controlling for variability within the groups. An effect size is an estimate of the magnitude of treatment effect.

Effectiveness The extent to which intervention results in favourable outcomes under everyday conditions. Contrasted with efficacy, which refers to change under tightly controlled conditions.

Expressive language Language produced by the speaker. In contrast to receptive language. See also syntax and morphology.

External validity The representativeness of the research findings and the extent to which it is possible to generalise the results.

Fluency The skill of using speech without undue repetitions or pauses.

Generalisability See external validity.

Generalisation The extent to which behaviours learned in one context can be transferred to another context or to other behaviours or stimuli (i.e. the use of trained behaviour in untrained situations).

Gold standard Term used for a standardised clinical assessment (or intervention) of known

continued

continued

validity and reliability which is generally taken to be one of the best available. Synonymous here with reference test, in the context of diagnostic testing.

Heterogeneity In research synthesis, heterogeneity refers to the variability of a collection of sizes of effects. Tests are available for the synthesist to check whether a given collection is more varied than would be expected on the basis of sampling variation alone.

Homogeneity A homogeneity test on effect sizes checks whether the effect sizes show greater variability than would be expected if their corresponding effect size parameters were identical.

Hybrid intervention approaches Combination of didactic and naturalistic approaches, such as milieu therapy.

Imitation Where a child is asked to repeat exactly what has been said by the adult.

Incidence The number of new cases of a disorder in a given time period.

Internal validity The extent to which research designs permit an interpretation regarding the causal relationship between an experimental treatment and an effect.

Intervention An explicit application of therapeutic/educational techniques intended to modify an individual's performance in a designated area associated with communication (i.e. expressive language, attention, etc.).

Language The set of symbols (usually words or signs) which are organised by convention to communicate ideas.

Learning difficulty/learning disability Generalised reduction in cognitive abilities, which usually impacts on language development (UK usage).

Longitudinal study Measuring the performance of a sample on more than one occasion over a period of time.

Meta-analysis The statistical analysis of the results of a collection of individual studies for the purpose of integrating their findings.

Modelling Where the child is asked to listen while the adult produces an example (model) of a target.

Monophasic Descriptive term for clinical assessments that assess one skill area only, such as speech/language, to the exclusion of other skill areas. See also multiphasic.

Morphology The part of grammar (syntax) which focuses on the components of words (roots, affixes, etc.).

Multiphasic Descriptive term for clinical assessments that assess more than one skill area, such as motor skills with language skills.

Mutism A condition in which the person does not use speech. This may be involuntary or by choice. If by choice, the condition may be generic (elective mutism) or restricted to certain situations (selective mutism).

Naturalistic intervention approaches Where the adult responds to the child's focus of attention rather than imposing a different context specifically for intervention.

Norm-referenced measure A score from a test that has been standardised on a population.

Neurodevelopmental Descriptive term for the maturing neurological systems of the child (e.g. motor skills, and perception).

Norm-referenced test An assessment measuring performance against standardised norms for chronological age (e.g. expresses performance score as an age equivalent). Compare criterion-referenced test.

Otitis media An infection of the middle ear, common in childhood. The effect of otitis media on language development has been much explored.

Parent–child interaction The joint engagement of parent and child in play or conversation.

Phonology The part of linguistic knowledge concerning speech sounds and their combinations.

PND Percentage of non-overlapping data – a technique for summarising data from single-subject experimental designs. It represents the percentage of treatment data points not overlapping previous baseline points.

Pragmatics The part of linguistic knowledge concerning use of language in social situations and the interpretation of communication contexts.

Prelinguistic skills Skills that are foundational for language development, including babbling, taking turns in simple play, eye contact, visual and auditory attention.

Prevalence The proportion of people in a population who have a given disease or attribute at a given point in time.

Prompting Where the child is encouraged to imitate or produce specific targets by means of questions or commands by the adult.

Prospective A study design that seeks to determine the association between a hypothesised risk factor and the occurrence of illness by sampling both exposed and unexposed subjects and then following them for the period of study. See also retrospective.

Quasi-experimental designs An experimental approach where subjects have not been randomly assigned to treatment and non-treatment groups, thus resulting in possible 'threats to validity' arising from differences in history of treatment, differences in maturation, and in biases in the selection of groups, which may reduce construct validity, internal validity and external validity.

Randomised controlled trial (RCT) An experimental approach wherein subjects from a sample, in this case, children with speech and language delay, are randomly assigned to either a treatment or a non-treatment group.

Receptive language Language heard and understood by a listener. Also referred to as verbal comprehension. In contrast to expressive language.

Reinforcement Where a correct performance by the child is rewarded by praise or a tangible reward.

Reference test See gold standard.

Reliability The degree of stability of measurement that exists when a measurement is made repeatedly under different conditions or by different observers.

Response generalisation Use of an untrained example at the same level of complexity as a trained example or a trained example at a higher level of complexity (e.g. in a sentence, rather than in a single word).

Retrospective A study design that seeks to determine the association between a hypothesised risk factor and occurrence of illness by sampling a group of subjects and investigating their prior exposure to the risk factor and their occurrence of illness. See also prospective.

ROC curve The receiver operating characteristic curve is a graphical representation of the pairs of true-positive and false-positive rates that correspond to each possible cut-off for the diagnostic test result. Used for ascertaining the optimum cut-off on a given screening measure.

Screening procedure Systematic procedure to select individuals from a given population at risk for an impairment.

Semantics The part of linguistic knowledge concerning the meaning of words and phrases.

Single-subject experimental designs An experimental approach involving the manipulation of an independent (treatment) variable across a pre-intervention 'baseline' phase, an intervention phase, and commonly a post-intervention phase, thus offering repeated measures of the outcome over time and providing data regarding the effects of the introduction, continuation, alteration and withdrawal of treatment. These designs can involve more than one subject.

Speech The physical production of language. See also articulation.

Speech and language delay Broad descriptive term for speech and language abilities which are considered to be below that expected for a child's chronological age, while still following the expected developmental sequence. Often qualified as mild, moderate or severe. For usage of the term 'delay' within this review, see page 1.

Speech and language disorder Broad descriptive term for speech and language abilities

continued

continued

which are considered to be develop-ing in a manner distinct from the usual developmental sequence. May be further qualified by noting those aspects of speech and language most affected: semantics, pragmatics, phonology, syntax.

Speech and language impairment General term for a speech and language problem, whether this is diagnosed as a delay or a disorder.

Stammering/stuttering Term used for hesitation, repetition or other disruptions in speech, which are considered to be outside the normal range of speech fluency. (American English usage: stuttering; British English usage: stammering).

Statistical conclusion validity Relating to the power and appropriateness of the statistical techniques used to analyse the data.

Stimulus generalisation The use of a learned response in a new setting, with new material or with unfamiliar people.

Syntax The part of linguistic knowledge concerning grammatical structures.

Treatment An explicit application of therapeutic/educational techniques intended to modify an individual's performance in a designated area associated with communication (i.e. expressive language, attention, etc.). See also intervention.

Validity Whether an experiment can explain what it claims to explain.

Note on WHO definitions

Impairment Dysfunction resulting from pathological changes in a system.

Disability Consequence of impairment in terms of functional performance (i.e. disturbance at the level of the person).

Handicap Disadvantages experienced by the individual as a result of impairment and disabilities. This reflects the interaction and adaptation to the individual's surroundings.

To these is often added the concept of distress or more recently well-being (Enderby 1992) reflecting the subjective response in the individual to the impairment.

References

The following sources were consulted for definitions; in addition, some glossary entries were supplied by the review team.

Cooper H, Hedges LV, editors. The handbook of research synthesis. New York: Russell Sage Foundation, 1994.

Centre for Reviews and Dissemination. Report number 4. Undertaking systematic reviews of research on effectiveness: CRD guidelines for those carrying out or commissioning reviews. York, UK: NHS Centre for Reviews and Dissemination, 1996.

Crystal D. The Cambridge encyclopaedia of language. Cambridge, UK: Cambridge University Press, 1987.

Enderby P. Outcome measures in speech therapy: impairment, disability, handicap and distress. *Health Trends* 1992;**24**:62–4.

List of abbreviations

AAPS	Arizona Articulation Proficiency Scale*
ACLC	Assessment of Children's Language Comprehension*
APP(-R)	Assessment of Phonological Processes (– Revised)*
ANOVA	analysis of variance
ANCOVA	analysis of covariance*
CA	chronological age
CBCL	Child Behaviour Checklist*
CDI	Communication Development Inventories*
CELF	Clinical Evaluation of Language Functions*
CELI	Carrow Elicited Language Inventory
DDST	Denver Developmental Screening Test*
DLS	Derbyshire Language Scheme*
DSS	Developmental Sentence Score*
DWSP	Developmental Word Sequencing Program*
EAT	Edinburgh Articulation Test*
EOWPVT	Expressive One Word Picture Vocabulary Test*
EPVT	English Picture Vocabulary Test*
GFTA	Goldman-Fristoe Test of Articulation*
GP	general practitioner
ITPA	Illinois Test of Psycholinguistic Abilities
LA	language age
LIPS	Leiter International Performance Scale*
LR	likelihood ratio
LT	late talker*
PND	percentage of non-overlapping data
MLU	mean length of utterance*
MSEL	Mullen Scales of Early Learning*
NSST	Northwestern Syntax Screening Test*
PCC	Percentage Consonants Correct*
PLS	Preschool Language Scale*
PPV	positive predictive value
PPVT(-R)	Peabody Picture Vocabulary Test (– Revised)*
RAPT	Renfrew Action Picture Test*
RCT	randomised controlled trial
RDLS	Reynell Developmental Language Scales*
REEL	Receptive Expressive Emergent Language Scale*
ROC	receiver operating characteristic
SD	standard deviation
SES	socio-economic status
SICD	Sequenced Inventory of Communication Development*
TACL	Test of Auditory Comprehension of Language
TOLD	Test of Language Development*
WILSTAAR	Ward Infant Language Screening Test, Assessment, Acceleration and Remediation

* Used in appendices only

Executive summary

Background

This report concerns the identification and treatment of children with primary speech and language delays, that is delays which cannot be attributed to other conditions such as hearing loss or other more general developmental disabilities. Such delays are important because they cause concern to parents, because they are commonly associated with behavioural and other difficulties in the pre-school period and because they constitute a risk factor for subsequent poor school performance, and for a wide range of personal and social difficulties for the individuals concerned. It is unclear, given the current state of knowledge, whether such delays represent varying levels of a single condition or a number of different conditions with diverse aetiologies.

Currently the identification and treatment of speech and language delays fall within the remit of the health services in the early years of life and most health trusts have in place informal procedures for identifying such delays. The educational services and those responsible for providing nursery and child-care services also have a considerable role to play in the process of identification and management of these children. This review aims to provide the information needed to help decide whether universal screening for speech and language delays should be implemented within the NHS.

Objectives

Four domains (prevalence, natural history, intervention and screening) were identified as being key to a review of screening issues, with the following objectives being stated:

- to undertake a systematic review of research into the value of screening and intervention for speech and language delays in children up to the age of 7 years
- to identify priority areas in need of further investigation
- to provide evidence-based direction for the future provision of services.

Methods

The review was carried out using structured guidelines for systematic reviews. These are described in detail in the full report.

Results

Prevalence

The number of potential cases of primary speech and language delay is high, with a median figure of 5.95% reported for delays in either speech or language. There has been little attempt to tie this evidence into prediction of subsequent case status, and there is little published evidence to support the perception that either the total number of children with language delay declines in real terms across the age range, or that prevalence has been rising over recent years.

Natural history

The natural history data indicate that a substantial proportion of children identified on the basis of expressive delay alone are likely to have difficulties which resolve spontaneously in the pre-school period. However, the data do not, at this stage, make it possible to predict **at the time of identification**, which of the children with **expressive** delay are likely to have persistent problems. A poorer prognosis has been consistently identified for children with expressive/receptive delays. The picture for older children is clouded by the lack of evidence from samples that have received no additional educational or therapeutic support. Nonetheless it is clear from follow-up studies of treated samples that children identified as having language difficulties in the first year of primary school are likely to have difficulties which persist through to secondary school.

Intervention

Results from randomised controlled trials (RCTs) and quasi-experimental designs reveal positive and statistically significant effects of intervention relative to untreated controls in all areas of speech and language skills. Comparable results for direct (clinician-administered) and indirect treatment were observed in the case of **expressive language**. In contrast, direct intervention was more effective in the case of **speech**, whereas indirect intervention

was more effective in the case of **receptive language**. Data from the single-subject experimental designs were synthesised and provide confirmatory evidence for the positive effects of intervention. The data in particular provide evidence for the generalisation of treatment effects. However, the data reviewed do not provide information about long-term outcomes of intervention, nor of the likelihood of intervention reducing prevalence in a given population. Similarly, it is not possible to draw conclusions about the effects of subject variables such as socio-economic status or age upon the relative value of interventions.

Screening

The screening evidence indicates that, although a considerable number of assessments have been shown to perform adequately in terms of their productivity, few studies compare the performance of two or more screening tests when applied to one population, nor do they compare single screening measures across different populations. It is difficult, therefore, to make judgements about the relative value of different procedures. In general, specificity is higher than sensitivity, suggesting that it is easier to determine who is not a case than to establish who is. Parent-focused measures appear to be as useful as specific tests of child behaviour. Interpretation is further complicated by the considerable variation in the cut-offs adopted on the range of reference 'gold-standard' measures, suggesting that there remains considerable disagreement as to what proportion of the population should be considered cases. There have been no explicit attempts to benchmark the target population in terms of prevalence estimates, the prediction of case status or the impact of the intervention.

Conclusions

It is clear that early speech and language delay should be a cause for concern to those involved with child health surveillance because of the problems for the individual child, because it may indicate other co-morbid conditions such as hearing loss, developmental and behavioural difficulties, and because of the implications it may have for literacy and socialisation in school. The fact that there is not sufficient evidence to merit the introduction of universal screening does not imply that speech and language delay should not be identified, for example, by less formal methods.

Implications for policy

The review suggests that more attention might be shown to the role of parents in identifying children with speech and language delay. Primary-care workers (health visitors, general practitioners, school nurses and nursery staff) should be involved in eliciting parental concerns and in making appropriate observations of children's communication behaviours. This would require formal training in delayed speech and language development and risk factors pertaining to it. Appropriate information would also have to be made available to parents to allow them to play an active role in judging need.

Given the reported value of indirect approaches to intervention there is a case for widening the range of professionals able to promote good interactive practice in parents of young children. Speech and language therapists as a professional group are in a good position to play an active role in disseminating this information and coordinating such services. Children who do not respond to such primary prevention could then be given access to speech and language therapy services and appropriately structured nursery input.

Recommendations for research

There are many gaps in the literature, and the review identified a number of research priorities.

- The impact of speech and language delay needs to be examined, both as an explanatory and a response variable across time in prospective cohort studies.
- RCTs need to be designed to examine the medium- and long-term effects of well described models of intervention. These should include an appropriate range of outcome measures including, where possible, economic analysis.
- There is a need for the development of a screening measure that combines data on risk factors with parental report and professional observation, and for the examination of its value in different sections of the population.
- The predictive ability of different models of early identification and intervention needs to be examined.

Further details of conclusions and recommendations are given in the full report.

Chapter 1
Background

The context

Primary and secondary delay

Speech and language delay is an umbrella term covering a range of conditions in early childhood. It is conventionally divided into **primary** and **secondary** delays. Primary delay occurs where the speech and language skills of the child are delayed relative to other skills, usually in the absence of a clear aetiology. Secondary delay occurs when the speech and language skills are delayed to the same extent as other skills, often as a result of a known aetiology, most commonly general learning disability or hearing loss. This report focuses on children with primary speech and language delay. It does not cover speech and language delays associated with other developmental conditions, of which speech and language delays are an associated symptom, such as autism, cerebral palsy, hearing loss, cleft palate, stammering and selective mutism. In addition, it does not cover medical conditions with which speech and language delays frequently co-occur such as primary psychiatric disturbance or high-risk neonatal histories.

The term **delay** suggests that it is possible to characterise this group of children along a single axis. In fact, speech and language represent a complex interaction of functions. Children may present both with different levels of delay and with qualitative differences in their presenting symptoms. For example, in some children speech alone is affected, for others problems may occur in expressive language and/or verbal comprehension. Some illustrations of typical cases of speech and language delay are provided in appendix 1. Reference is sometimes made to a distinction between **delayed** speech and language, which follows the normal pattern and has a more benign outcome, and **deviant** or **disordered** speech and language, which does not follow the normal pattern and has a more adverse outcome. Recent developments have suggested that there is not a clear distinction between delay and disorder (Curtiss *et al,* 1992). Some authors have employed the term speech and language **impairment** suggesting the constitutional nature of the problem. The term **specific speech and language impairment/disorder** is also used. This is akin to primary speech and language delay but it tends to be defined in terms of a discrepancy between verbal and non-verbal skills measured on standardised assessments. This practice of defining the specific condition using discrepancy scores has been called into question but remains popular in the research literature (Aram *et al,* 1992). The term speech and language delay will be retained throughout this report.

The review covers children with speech and language delays in the 0–7 years age range. This is the period of primary clinical interest as far as health services in the UK are concerned because children are most commonly referred to speech and language therapy services before they reach school (The Department for Education, 1994). It also corresponds to the period of most active linguistic development and is the period which has, to date, received the most attention from the various professional groups involved with these children.

There is a wide range of variation in both the speed and the quality of language acquisition in the early years. Yet for some children this variation constitutes a delay in the acquisition of speech and language skills such that it affects performance and warrants concern from those in the child's environment. The concern is that such delays adversely affect the child's ability to develop at both a personal and a social level and that the difficulties may lead to disadvantages in terms of educational performance and subsequent social development. There has been considerable discussion about the nature of the underlying difficulties experienced by children with primary language delays. Much of this revolves around whether it is possible to identify a specific intrinsic mechanism or series of mechanisms common to all such children or whether it is more appropriate to look for extrinsic mechanisms. Some authors have posited a highly specific auditory processing disorder (Tallal and Piercy, 1973; Wright *et al,* 1997) that underlies both specific language delays of this type and dyslexia. Others have questioned this evidence (Studdert-Kennedy and Mody, 1995) and raised other possible explanations (Bishop, 1992). Epidemiological evidence presents rather a different picture pointing to the high levels of co-morbidity associated with speech and language delay. These children often experience other conditions that are

relatively common in childhood, such as neurodevelopmental delays, poor attachment to the primary care-giver, behavioural and psychiatric problems and otitis media (Bax *et al,* 1990), and some authors have suggested that a more general neurodevelopmental delay may account for most speech and language delays in the young child (Stevenson, 1996). Despite this level of association these speech and language delays may still be considered primary rather than secondary because the most readily identifiable difficulty is of speech and language. To this extent primary speech and language delay in the young child may be taken as a societal construct. As the 'noise' of normal development recedes with advancing age specific mechanisms may become more apparent.

Factors affecting speech and language development

The extent to which delays in speech and language development may be attributed to external factors, such as social class, has attracted much attention. The so-called 'verbal deprivation theory' suggested that children presented with poor language skills because they received inadequate models from their environment. This interpretation is now widely discredited as an explanation of differences in speech and language performance (Tizard and Hughes, 1984; Wells, 1985). However, many children of families from lower socio-economic groups present with a marked difference in performance between home and school/nursery environments, the conclusion being not that they do not have the necessary language skills but that they do not display them in less familiar contexts. Clearly this is an area of concern for any screening programme. Heavy reliance on the decontextualised measurement of skills may increase the child's chances of being inappropriately identified as delayed.

Furthermore parents are variable in the extent to which they respond to their child's communication attempts (Wells, 1985). For the purposes of the discussion here the central issue is the degree to which parents are able to support the speech and language needs of children who are slow to develop their speech and language skills. Children with difficulties processing and formulating language need responsive and **available** communication partners to help reinforce their attempts to communicate. Lack of availability of a communication partner and, in particular, lack of opportunity to interact (especially when taken in the context of familial stress) are much more likely to be determinants of whether a child's speech and language can be optimised, than socio-economic status (SES) itself.

The importance of stimulation and the role played by linguistic input highlights the needs of one group of children for whom the language environment is particularly distinctive, notably those exposed to more than one language. Concern has been expressed regarding the accurate identification of bilingual children who may have delayed language development. Although children brought up in bilingual environments often experience periods in which they mix languages there is no evidence that bilingualism as such should be considered a contributor to clinical levels of language delay or indeed that language delays occur more frequently in bilingual populations (Genesee, 1988). Indeed it seems likely that, for many children, bilingualism is an advantage in the insight that it can offer into the structure of language. Nonetheless, bilingualism remains a complex issue for those with a responsibility for identifying children with speech and language delays. This is particularly so where there is inadequate support from bilingual services and where little is known about the other languages to which the child is exposed.

The role of health and educational services

Speech and language delays fall within the remit of the health services in the early years for both conceptual and pragmatic reasons. At a conceptual level, speech and language reflects the health and well-being of the child both in terms of the child's capacity to externalise his or her needs and in the role it plays in mediating the child's internal states. At a more pragmatic level, those with the initial responsibility for the identification of children with speech and language delay, namely general practitioners (GPs), health visitors, clinical medical officers and community paediatricians are an integral part of the health services. Of these, health visitors are the principal source of referrals of children within the pre-school period (Jowett and Evans, 1996). Speech and language therapy services, which play a large part in the treatment of these conditions, are funded by the health services in the UK, though educational services also play a major part in funding therapy in Scotland. Speech and language development is also intimately related to all aspects of educational and social development and, as such, delays in these areas also fall within the remit of educational services and of those responsible for providing nursery and childcare services.

The need to examine the research relevant to screening for speech and language delays must be seen against the backdrop of a perceived increase in demand for services for these children. Although the process of systematic identification of children

with developmental delays began in the 1950s, it was consolidated in the UK by the recommendations of the report *Fit for the future* (Court, 1979), and by the recognition that a number of these children go on to have special educational needs (Warnock, 1978). Local services then developed across the UK to meet the needs of children with developmental delays, but this has largely been in the absence of any centralised policy governing the approach. No central records are kept of the number of children with speech and language delay across the UK. However, the increase in demand for services for these children in recent years is reported to be of the order of 30% (Jowett and Evans, 1996) with 40% of all children being in the 0–4 years age band (Reid *et al,* 1996). Figures for services to all children with speech and language delays within the educational services are collected locally and are not available at a national level. The Scottish Office has recently produced a report indicating that the number of children with Records of Needs for language and communication disorders had risen from 179 to 363 between the years 1989 and 1993 (The Scottish Office: Education and Industry Department, 1996). These figures only reflect those children with the most persistent or severe problems and consequently mask the real level of potential need. Inevitably this increase in demand has a knock-on effect on the service providers. However, as indicated above, children with speech and language delays are not the sole preserve of the speech and language therapy profession and this increase in the number of perceived cases will have an effect on the demand for provision offered by medical and educational services.

Prevention

Whether it is possible to have an impact on such figures depends on the extent to which it is possible to prevent speech and language delays. Prevention in health terms is usually divided into three components (Butler, 1989).

- **Primary prevention** is aimed at the promotion of good health by reducing the incidence of disease and other departures from good health.
- **Secondary prevention** is aimed at reducing the prevalence of disease or departures from good health by shortening their duration or diminishing their impact through early detection and prompt and effective intervention.
- **Tertiary prevention** is aimed at reducing impairments and disabilities, minimising the suffering caused by existing departures from good health and promoting the child's adjustment to conditions that cannot be ameliorated.

It is possible to fit speech and language delays into this prevention paradigm. Speech and language delays may be construed as departures from good health to the extent that speech and language skills reflect the child's well-being. This review will be primarily concerned with the feasibility of secondary prevention of which screening is an integral component.

Screening

A number of criteria have been proposed by those working within the medical model that need to be met prior to the introduction of universal screening (Wilson and Jungner, 1968).

- The condition should be an important health problem.
- The natural history of the condition should be known.
- The condition should have a recognisable pre-symptomatic stage.
- There should be an acceptable and effective form of treatment at the pre-symptomatic stage.
- There should be an agreed policy upon whom to treat.
- There should be facilities for investigation and diagnosis.
- There should be a screening test available.

There has been some resistance to the introduction of screening for all but the most clearly defined medical conditions on the grounds that there is relatively little evidence that there are demonstrable differences in later life between children who have and have not participated in surveillance programmes at pre-school ages (Butler, 1989). Whether speech and language delays fit readily into such a medical framework is something of a moot point. Indeed, there are a number of characteristics (listed below) of speech and language development in general, and delays in particular, which suggest that the above criteria may be not be appropriate.

- Speech and language development is multidimensional. It includes aspects of speech, vocabulary, syntax, morphology, verbal comprehension, etc. Screening and indeed intervention will, therefore, be operating on subsets of symptoms rather than a single entity.
- A child's developmental status is continually changing over time and judgements about strengths and weaknesses may be dependent upon expectations. Similarly, behaviours may be adaptive or maladaptive according to context. Two children with similar presenting features may react differently to their environments, one being seen as problematic, the other as coping.

- The issue of natural history is complicated by the changing nature of the conditions. Persistence of the speech and language difficulties themselves are likely to be important in their own right, but such difficulties may also be associated with social, educational and behavioural difficulties in the later years. Thus, there may be no clear linear relationship between initial presentation and subsequent outcome (Lichtenstein and Ireton, 1984).
- It is debatable whether it is possible to identify a pre-symptomatic stage of a developmental condition. It may be possible to establish that a child's speech and language skills are delayed at a given point but whether all delays thus identified are necessarily cases is somewhat uncertain. It may be more acceptable to refer to the identification process as being at a pre-diagnostic rather than a pre-symptomatic stage.
- Case definition may be carried out in a number of different ways. It is possible to specify cut-off scores on standardised tests of speech or language development. But this alone is not sufficient to determine whether a child needs to be treated. To date there has been relatively little research examining the process of how clinicians make their clinical judgements. It is likely that case status is rooted in the perception of a child's performance on criterion- or norm-referenced procedures rather than the extent to which it is known that a child can be treated effectively. For example, using clinical profiles, Records and Tomblin (1994) found that the probability of positive diagnosis increased as the tested level fell more than –1.2 standard deviations (SDs) below the mean. Although this is interesting in so far as it suggests that case status reflects the normal distribution, it is also subject to circularity in that the clinicians made use of test results in their judgements and this, in turn, may reflect their training regarding the psychometric properties of assessments.
- The highly social nature of speech and language delays also raises the question of the role played by others in the definition of case status. Parents, teachers and care staff are likely to be in a position to observe the extent to which the child's level of skill affects his or her ability to communicate effectively.
- Approaches to intervention vary considerably. In this review they have been broadly grouped according to the method of delivery (see appendix 2 for examples), but it is recognised that children with speech and language delay often have therapy tailored to their individual needs. The effect of the individual skills of speech and language therapists and teachers is also an issue which has remained largely unexplored.
- Although it is true that identification of a child with a developmental delay may lead to resources being directed towards that child, it is not true that such a child will receive no help unless that identification is made through the process of screening. The educational environment provides a level of input that will have an impact on the child's developmental status irrespective of whether he or she has been defined as a case or not in a screening programme.
- The criteria above refer to the evidence required **prior** to the introduction of universal screening. The question of whether the same evidence could be used to argue for the **withdrawal** of existing services is much less clear. This point will be discussed further below.

The four domains of enquiry

In order to address the question of whether there is a case for speech and language delay to be included in a screening programme, this review will examine four domains of enquiry:

- the prevalence of speech and language delays
- the natural history of speech and language delays
- the effectiveness of intervention approaches for speech and language delays
- the accuracy of screening procedures.

An additional set of questions has been generated for each of the four domains in order to structure the discussion in each section. These questions have been derived from the review team's knowledge of the literature and clinical practice in speech and language therapy and educational psychology.

The structure of a screening model

It is useful to construct a model of the detection and treatment process to indicate what information would be needed to examine properly the feasibility of screening for speech and language delay, and to place in context the additional questions addressed in chapters 3–6. Such a model is given in *Figure 1* and shows two populations being compared. The upper half of the figure represents a population in which children are routinely screened for speech and language

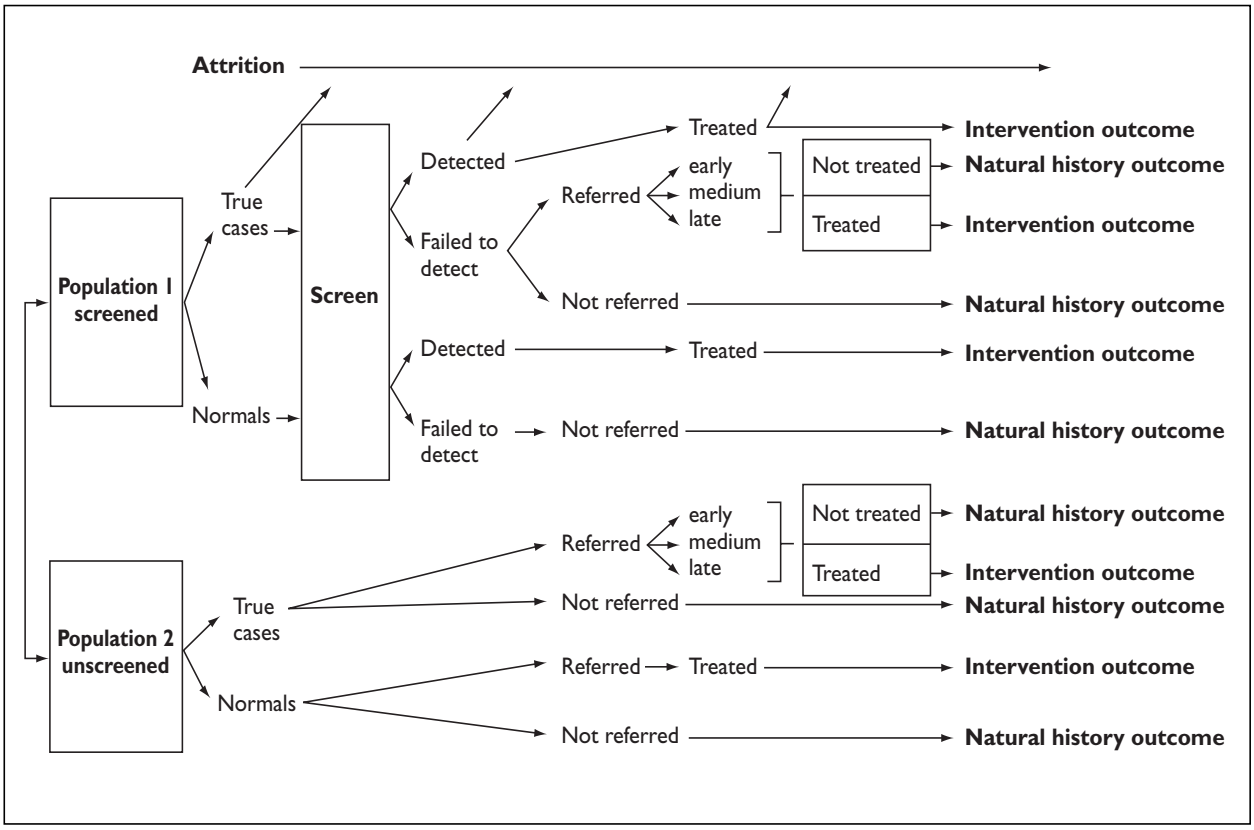


FIGURE 1 *A model of the detection and treatment process*

delays. The lower half of the figure represents a population for whom no such formal screening takes place. In the first instance it would be of value to compare these two populations for uptake of services and outcomes following intervention. The proportion of true cases in a population reflects the **prevalence** of speech and language delays that the screening procedure is attempting to capture. The number of true cases identified by **screening** depends upon the accuracy of the measure in question. The consequences of screening can be expressed in terms of screen-positives and screen-negatives, and then in terms of whether they have been correctly detected (i.e. are true- or false-positives). Those true cases not picked up by the screen (false-negatives) may nonetheless be picked up by other referral mechanisms. True-positives are depicted in terms of whether or not they go on to receive intervention, in which case they provide **treatment outcomes**. If they do not receive intervention having been identified they provide evidence of **natural history** or the relative impact of intervention. One of the greatest threats to the validity of a screening programme is the extent to which it effectively covers the population in question. At each stage of the process there is likely to be a drop in coverage due to, for example, poor attendance or lost contacts. Such attrition of the screened population needs to be monitored throughout.

The figure suggests that it should be possible to read through from left to right, from prevalence through to treatment outcomes. However, it is also possible, and this is broadly the position taken here, to construe this table from right to left, such that the evidence from natural history and intervention studies allows the identification of a group for whom intervention can be shown to work and who would not otherwise improve spontaneously. Once this group is accurately described it should be possible to screen the population for the group that stands to benefit from intervention in order to establish its prevalence.

Chapter 2
Methods

The NHS Centre for Reviews and Dissemination's *Guidelines on undertaking systematic reviews* (1996) were followed throughout. The review team were also assisted throughout by a multi-disciplinary advisory group of researchers and practitioners and other specialists.

Search strategy

A preliminary literature search of the following systematic review databases failed to locate any similar reviews completed or in progress:

- Cochrane Database of Systematic Reviews
- Database of Abstracts of Reviews of Effectiveness
- The Cochrane Controlled Trials Register
- The Cochrane Review Methodology Database.

Literature databases

The databases most likely to yield relevant literature were selected on the basis of a scoping search using the CROS and DialIndex search mechanisms (see appendix 3). Databases that included unpublished work were also checked. The following databases were then selected for development of comprehensive search strategies:

- Cumulative Index of Nursing and Allied Health
- EMBASE
- Educational Resources International Clearing House
- Linguistics and Language Behaviour Abstracts
- MEDLINE
- PsychLit.

The search strategies developed were designed for maximum retrieval, using both indexing terms and free text searching.

Handsearches of journals

The following journals were also searched by hand (missing issues are noted in parentheses):

- *Australian Journal of Human Communication Disorders* 1986–1995
- *British Journal of Disorders of Communication* [continued as *European Journal of Disorders of Communication*] 1986–1995
- *Bulletin of Royal College of Speech and Language Therapists* 1986–October 1996 ('86: Feb; '87: May; '88: Feb, Dec; '89: Jan–July, Oct, Dec; '90: Feb–Dec; '91: Jan, April–Sept, Nov, Dec; '92: Jan–May)
- *Child Language, Teaching and Therapy* 1986–1995
- *Journal of Child Psychology and Psychiatry and Allied Disciplines* 1990–1996:5
- *Journal of Speech and Hearing Disorders* 1980–1990
- *Journal of Speech and Hearing Research* 1986–1995
- *Health Visitor* 1990–1996:1 ('92:9; '93:1; '94:11; '95:8)
- *Language, Speech and Hearing Services in Schools* 1986–1995
- *Monographs of the Society for Research in Child Development* 1986–1995
- *Special Education* 1986–1995.

Compilation volumes

The following compilation volumes were searched:

- Yearbook of speech, language and hearing (1990 and 1991). Bernthal JE, Hall JW, Tomblin JB, editors. St Louis: Mosby Year Book, Inc. (Critical compilation of key articles drawn from 67 international journals [series then discontinued].)
- Enderby P, Emerson J. Does speech and language therapy work? London: Whurr Publishers, 1996.
- Index to recent literature on speech and language (September 1996–January 1997). C Norris. Biomedical Research Indexing.

Bibliographies

The following bibliographies were searched:

- the bibliography from Hall DMB. Health for all children. 3rd ed. Oxford: Oxford University Press, 1996. (Supplied on disc by the College of Paediatrics and Child Health, 5 St. Andrew's Place, NW1 4LB.)
- a bibliography on screening issues, supplied by Dr H Ireton (personal compilation)
- The American Speech Language Hearing Association. Treatment efficacy bibliography.

American Speech Language Hearing Association, 1995
- reference lists taken from the articles retrieved.

Internet sources (WWW, TELNET)

An experimental meta-search engine, *Savvysearch* (http://guaraldi.cs.colostate.edu:2000/), was first used to query multiple Internet search engines simultaneously. This identified the most relevant Internet search engines (including speech and language issues) as:

- Webcrawler (http://webcrawler.com/)
- Infoseek (http://www2.infoseek.com/)
- Excite (http://www.excite.com/)
- Yahoo (http://www.yahoo.com/)

Other useful addresses were obtained from the NHS Centre for Reviews and Dissemination Information Service:

- The National library of Medicines' Health Services/Technology Assessment Text (http://text.nlm.nih.gov/)
- Research Activities and Publications Information Database (http://edina.ed.ac.uk/rapid/) (NB: this search required TELNET access [ercvax.ed.ac.uk]).

Calls for information to Internet users resulted in the identification of the following relevant addresses:

- Speech Pathology WWW Sites (http://www.glasswing.com/~shh/speech.html)
- Amazon.com Books (http://amazon.com)

Unpublished literature

Two databases dealing with unpublished literature were scanned for any relevant material (see appendix 3 for search terms used):

- System for Indexing Grey Literature in Europe
- Boston Spa Conferences (British Library database)

Calls for information

Calls for information were made to professional organisations, institutions and authors (see appendix 3). Literature identified in this way was also incorporated into the decision-making process.

Additional information

Information was brought to the attention of the team throughout the review. It was necessary to set an end date for retrieval and coding of articles (9 May, 1997). Thus some material from outside the UK which could not be located before this date, and papers arriving late were not included. This has been noted where possible. A total of 9983 papers were retrieved across the four domains of prevalence, natural history, intervention and screening. After automatic de-duplication, the references retrieved were sorted, together with other literature, according to the inclusion/exclusion criteria (see appendix 4).

Inclusion criteria

Judgements regarding whether studies retrieved should be included or excluded from this review were based upon both the relevance of the study and its acceptability, taking into consideration the design of the study and protection against threats to:

- **construct validity** (the ability of a measure or an intervention to reflect an underlying construct)
- **internal validity** (the extent to which research designs permit an interpretation regarding the causal relationship between an experimental treatment and an effect)
- **external validity** (the representativeness of the research findings and the extent to which it is possible to generalise the results)
- **statistical conclusion validity** (relating to the power and appropriateness of the statistical techniques used to analyse the data).

(The inclusion/exclusion criteria for each domain are presented in appendix 4.) Studies which did not meet the criteria for relevance (i.e. on grounds of date of publication, age of children, nature of language delay, or falling outside the stated range) were not logged. Relevant papers which failed on aspects of acceptability (i.e. validity) were logged as 'excluded papers' without being included in the data extraction stage. Details of the reason for rejecting such excluded papers were logged using the information from the appropriate inclusion/exclusion table. Studies that met the inclusion criteria are summarised in appendix 5. Final judgements about inclusion were made by two independent assessors, with discussion to resolve any disagreements (see appendix 6 for those studies that were excluded and reasons for exclusion).

Data extraction

Studies which met the relevant criteria were coded using the appropriate data extraction form for each domain (see appendices 7 and 8). The data

extraction forms were designed to minimise coding errors by preserving as much of the original information from studies as possible, thus reducing the number of judgements that the coder was required to make. See appendix 7 for reliability of coding.

Where possible, non-numerical information was coded into a numerical format, using lists of fixed and well-defined categories. The categories used provided a means of testing hypotheses and were selected on the basis of (a) knowledge of the literature and (b) likely effect modifiers. Additional categories were added as necessary during the process of data extraction and coding.

External validity

Where possible, checks were made to ascertain the representativeness of the data. These included consideration of the effects of publication bias and of Type I and Type II statistical error. In the case of the intervention data, the following checks on the possible effects of Type I error (i.e. that the null hypothesis might be erroneously rejected and positive results from intervention incorrectly assumed because of unrepresentative data) were carried out:

- the use of 'funnel plots' of standardised effect size by sample size and by study quality to examine whether there were gaps in the data set due to publication bias in favour of significant results
- inspection of the direction of outcomes from excluded studies to examine the extent to which studies excluded on grounds of design also showed positive effects of intervention.

Data synthesis

Numerical data from the four domains were pooled across studies as shown in *Table 1*.

In addition to the pooling of quantitative data in this way, non-numerical data were analysed using qualitative approaches, where appropriate. Full details of the data synthesis can be found in the chapters that follow.

TABLE 1 *Basis for data synthesis of numerical data in the four domains of the review*

Domain	Data synthesis
Prevalence	Medians of prevalence rates across studies (summarised by age and type of impairment)
Natural history	Medians of the percentage of children across studies whose speech and language delays resolved without treatment
Intervention	Standardised effect sizes for the outcomes from randomised controlled trials (RCTs) and quasi-experimental designs and the PND statistic (percentage of non-overlapping data between baseline and post-baseline phases) for single-subject experimental designs
Screening	Sensitivity and specificity rates and likelihood ratios (LRs)

Chapter 3

The prevalence of speech and language delay

Review questions

- What is the reported range of prevalence estimates?
- What evidence is there of different prevalence rates for different subgroups of speech and language delays (e.g. delays of expression only, of expression plus comprehension)?
- How is it possible to account for differences within this range (e.g. identification procedure, age, gender, SES, sample, cut-offs)?
- What implications do estimates of prevalence have for the screening process?

Inclusion/exclusion of literature

Relevance	Studies of prevalence of speech and/or language delays in children up to 16 years.
Participants	Information about the number of participants in the population and the diagnostic samples.
Outcomes	Standardised measures of speech and/or language and clearly defined clinical judgement.
Designs	(a) Subsample of a normal population (b) Normal population screened and then sampled (c) Complete normal populations.

Studies were included that estimated the prevalence of speech and language delays in children aged up to 16 years in a general population (see appendix 4). Those studies with samples taken from clinical populations were excluded. In addition, a distinction was drawn between single-level prevalence studies, which have been based on surveys or unvalidated clinical judgement, and those that have been based on a pre-screen or net of the population with proportions of passes and fails sampled and then given a diagnostic assessment, either on a standardised language procedure or on a criterion-referenced clinical judgement (which the authors have made some attempt to validate or define such that it could be replicable). The former have been excluded on the grounds that they would be impossible to replicate and are likely to lead to under-reporting in all but the most clear-cut medical cases (Leske, 1981). The largest studies have derived figures from household questionnaires; generally these have also reported the lowest prevalence figures (Blum-Harasty and Rosenthal, 1992).

The advantage of the approach that relies on diagnostic testing is that the resulting data are easier to compare with those from other studies. The disadvantage is that there is an implicit circularity in setting a cut-off on a measure that has been developed on a normal population. By definition, that cut-off will imply a prevalence rate. Any discrepancy between the implied prevalence rate and that found in a prevalence study can better be explained by the potential differences between the two populations than by differences in the rate of true cases in the population studied. Cut-offs tend to follow psychometric convention without any direct attempt to link them with clinical judgement. Clinical judgement of case status is also prone to circularity because it is likely to relate to availability of services and to the expected response to therapy. Stronger support for prevalence levels would be indicated if the rates derived by clinical judgement of cases showed agreement with those rates derived using conventional cut-off scores.

Review of the data

Since the publication of *A note on the prevalence of speech and language disorders* (Rutter and Martin, 1972), there have been a number of attempts to draw together the literature on the prevalence of speech and language delays. In most cases they are just that – lists of prevalence figures based on a variety of designs used to access the population as a whole and based on a variety of methods for ascertaining the skills of the children concerned. In order to represent its diversity the data in the present review were classified by the language domains measured, and by the age of the children in the sample. *Table 2* shows prevalence for speech and language delay and *Table 3* summarises studies of expressive and receptive delays where data on speech difficulties were not recorded. *Table 4* shows studies that report period prevalence, which by definition are impossible to combine in a meaningful way. The summary tables for included prevalence studies are provided in appendix 5.

TABLE 2 *Median prevalence estimates by type of speech and language delay and age*

Age (years;months)	Speech and/or language delay (median % [range])	Language delay only (median % [range])	Speech delay only (median % [range])
2	5.00[1]	16.00 [8.00–19.00][2]	–
3	6.90 [5.60–8.00][1,3,4]	2.63 [2.27–7.60][5,6,7]	–
4;6	5.00[1]	–	–
5	11.78 [4.56–19.00][8*]	6.80 [2.14–10.40][5,8,9,10,11]	7.80 [6.40–24.60][8,9,10]
6	–	5.50[9]	14.55 [12.60–16.50][9,10]
7	–	3.10 [2.02–8.40][5,9]	2.30[9]

Note: In a number of studies more than one data set is provided for a given age band. In order to avoid over-representing such studies only a single (median) figure was included for each study.

[1] Bax et al *(1980); [2] Rescorla* et al *(1993); [3] Randall* et al *(1974); [4] Burden* et al *(1996); [5] Silva* et al *(1983); [6] Stevenson and Richman (1976); [7] Wong (1992); [8] Beitchman* et al *(1986); [9] Dudley and Delage (1980); [10] Tuomi and Ivanoff (1977); [11] Tomblin (1997).*

[*] Beitchman et al *(1986) is the only study to include prevalence estimates for both speech **and** language, and speech **or** language. This highlights the difficulty in synthesising data in this area because not only is the level of difficulty not clear but the extent to which the categories can reasonably be teased apart is unclear.*

TABLE 3 *Median prevalence estimates of language delay by age, in the absence of speech delay*

Age (years)	Expressive and receptive language delay (median % [range])	Expressive delay only (median % [range])	Receptive delay only (median %)
2	–	16.00[1] [8.00–19.00]	–
3	3.01[2,3] [2.63–3.40]	2.30[2,4] [2.27–2.34]	2.63[2]
5	2.14[2]	4.27[2]	3.95[2]
7	2.02[2]	2.81[2]	3.59[2]

Note: In a number of studies more than one data set is provided for a given age band. In order to avoid over-representing such studies only a single (median) figure was included for each study.

[1] Rescorla et al *(1993); [2] Silva* et al *(1983); [3] Wong (1992); [4] Stevenson and Richman (1976).*

TABLE 4 *Prevalence studies covering different age ranges*

Age (years)	Study	Type of delay	Total (%)
2–9	Paul *et al* (1992)	Speech and language	1.35
6–12	Harasty and Reid (1994)	Speech and language	8.00
3–5	Stewart *et al* (1986)	Speech	1.50
5–7	Kirkpatrick and Ward (1984)	Speech	4.60
6–12	Harasty and Reid (1994)	Speech	12.60
12–14	Warr Leeper *et al* (1979)	Speech	7.30

Note: In a number of studies more than one data set is provided for a given age band. In order to avoid over-representing such studies only a single (median) figure was included for each study.

The most important group to characterise in terms of the feasibility of screening is the group with both speech and language difficulties. In this group the median prevalence is 5.9% (the median of the speech and/or language estimates, *Table 2*). However, considerable caution needs to be taken in extrapolating from this type of data synthesis to produce single composite prevalence estimates. The most obvious characteristic of these data is their variability.

Prevalence in subgroups of speech and language delay

Taking speech and language delay as a single construct, the majority of studies took diagnostic assessment score cut-offs between –2 and –1.5 SDs below the mean for the standardisation sample on the measure used, which automatically gives prevalence rates between 2.28% and 6.68%. In most cases the findings are higher than would be anticipated on this basis, ranging from 1.35%

to 19% (*Tables 2* and *4*). In the one study with a prevalence rate of 19% the more liberal cut-off of –1 SD was used. The study by Paul and co-workers (1992) reported much lower figures (1.35% for mild, moderate and severe cases, 0.65% for moderate and severe or 'serious' cases only). These figures derive from a study in which case status depended upon an objectified criteria for clinical judgement rather than performance on a standardised assessment of speech and language performance. Such conservative criteria may reflect the fact that this study was carried out in a developing country with relatively few resources for helping children with speech and language delay.

Some studies have reported children with either expressive or receptive language delays as one group (*Tables 2* and *3*). The variability here is also wide, giving a range of 2.02% to 19.00%. In one case, a cut-off that would anticipate a higher prevalence figure (–1.25 SD) resulted in a convergent prevalence estimate of 7.4% because children had to receive two low test scores to achieve case status (Tomblin, 1997). The figures for expressive and receptive language delay show little variation within a narrow band, reflecting the conservative cut-off scores adopted in these studies.

The figures for speech delay only are also variable, ranging from 2.3% to 24.6% (*Tables 2* and *4*). It is of interest that speech delay has not generally been a subject of prevalence studies in the pre-school years. The figures given by Warr-Leeper and co-workers, Stewart and co-workers, and Tuomi and Ivanoff were not based on specified cut-offs, and along with the Harasty and Reed study received relatively low quality rankings (see appendix 8). Kirkpatrick and Ward (1984) identified 4.6% using a –2 SD cut-off. Similarly, Beitchman and co-workers (1986) set a –2 SD cut-off on particular sub-tests to reach their speech prevalence rate of 6.4%.

Three studies reported figures for delays in expressive language skills only. Two of them used an expressive language measure only (Rescorla *et al,* 1993; Stevenson and Richman, 1976), while the other (Silva *et al,* 1983) sought to identify expressive delay in the absence of delay in receptive language skills. It is evident that some children failing an expressive language measure would also have receptive language difficulties. The figures reported by Stevenson and Richman, and Silva and co-workers ranged from 2.34% to 4.27% over the ages of 3–7 years. The one outlier study is that by Rescorla and co-workers (1993), with figures of 8%, 16%, and 19% according to the screen cut-off. These figures were computed from a single vocabulary checklist rather than a diagnostic test performance, which was the approach adopted in the other studies (i.e. the range of reported prevalence reflects the screening stage of this study). Alternative prevalence figures can be computed reflecting the reference measure, giving 9.8% and 13% (see appendix 5). Finally the age range for Rescorla's study was at least a year below that of the other studies, and it may be that the range of expressive vocabulary development is particularly wide in the slightly younger age group.

The figures for receptive delay are again tightly grouped ranging from 2.63% to 3.95%. All of these figures come from the Dunedin study (Silva *et al,* 1983). The prevalence of receptive delay in the absence of associated expressive delay is much more pronounced than would be predicted from the evidence provided in an earlier clinical study, which suggested that only two out of 74 delayed children had isolated receptive delays (Morley, 1965). The children in the Silva study are unlikely to meet the criterion for Morley's diagnosis. Silva's figures here could include children who are actually expressive-receptive language delayed, but who just achieved a pass on the expressive language measure when failing the receptive test. The figures may therefore be an instance of psychometric convention for cut-offs confounding the clinical impression.

Accounting for variable prevalence rates

Cut-offs

With the exception of Stevenson and Richman (1976) who obtained a prevalence of 0.6% for specific language impairment in the absence of other developmental difficulties, none of the above studies investigated the role of overall intelligence. This factor is likely to be implicated in a substantial number of speech and language delays. Between 2% and 3% of all the cases identified in these studies are likely to fall within conventional criteria for more general learning disabilities (MacKeith and Rutter, 1972). As with the level of speech or language functioning, levels of intellectual handicap are determined by arbitrary cut-offs on standardised procedures and the correlation between language and other intellectual skills is likely to remain relatively high, particularly when both expressive and receptive skills are implicated. Relying on standardised assessments to establish case status makes it difficult to judge whether prevalence decreases with age. A given cut-off on such a measure will result in the same percentage of the population being identified at any given time. The data presented here suggest that prevalence does not decrease over time. The study by Bax and co-workers (1983) reported fairly stable figures over

the age range 2;0–4;6 years for their 'definitely abnormal' group. It seems likely that the majority of studies reviewed above take this 'stable' group as their subject pool. This could suggest that similar prevalence figures across a time range reflect the same group of children at different points in time. In fact, the evidence from Silva and co-workers (1983) suggests that children may move in and out of the delayed group, though this raises questions about the test–retest reliability of the measures used. Predictability of problems depends upon the presenting symptoms at any one point, children with expressive and receptive delays being much more likely to have persistent problems than children with expressive delays alone. By contrast, Bax and co-worker's 'possibly abnormal' group shows a decreasing prevalence from 17% to 7% over the same time frame. This suggests that there is a group of children for whom development is particularly variable in rate but who may have less entrenched problems. These may tend to reduce over time perhaps for no other reason than test–retest error or regression to the mean. But a further possibility is that this reduction is a result of the effect of speech and language therapy services, a point made by Butler (1989) with regard to the Bax study.

Gender

For many years it has been clinically recognised that marked speech and language delays are more common in males than females and indeed this is generally confirmed by the studies reviewed here. Gender ratios quoted are 1.25:1 (Randall *et al,* 1974), 2.26:1 (Stevenson and Richman, 1976), 2.30:1 (Burden *et al,* 1996), 1.25:1 for both speech and language at 4 years (Stewart *et al,* 1986), and 2.3:1 (speech) and 1.2–1.6:1 (language) (Tuomi and Ivanoff, 1977). There are two exceptions to this pattern: Beitchman and co-workers (1986) found the reverse pattern for speech only (0.98:1), language only (0.98:1), and speech or language (0.82:1), and 0.46:1 for the speech and language diagnosis; and Tomblin (1997) suggested that, while boys are more likely to present with specific language impairment, the ratio is near equivalence. There are two possible explanations for these figures. The first is suggested by the design of the Beitchman study, which sought to sample and then project the false-negatives back into the original population sample. Of the false-negatives, the majority were girls and in projecting back up to the main sample the authors projected the gender balance as well as the number of cases. The other explanation (which, given the corresponding finding in the Tomblin paper is probably the more likely of the two explanations), is that the relatively liberal cut-off effectively misses the commonly observed discrepancy between the genders because those cases found may be less likely to be true clinical cases and as such may tend to reflect the normal gender balance in the population.

Socio-economic status

The studies quoted here are not helpful in addressing the issue of increased prevalence in lower socio-economic groups. A number of studies have commented on this issue but those that have met the inclusion criteria for this study have often deliberately excluded groups of people who might be considered of lower SES. Some of the studies were carried out in areas with a relatively advantaged population (Burden, 1996; Rescorla, 1993). Bax and co-workers (1983) commented on the cumulative effect of low SES on language delay. Similarly Harasty and Reed (1994) gave some indication of the potential effect of introducing variation of this type into prevalence estimates. Their study is of particular interest in that the authors adopted delay criteria very similar to Beitchman and his colleagues.

Language background

The data here do not address bilingual or ethnically diverse populations. Stevenson and Richman (1976) deliberately excluded non-indigenous families. Tomblin (1997) noted increased prevalence of language delays in monolingual African Americans, a finding not replicated by Stewart (1986) in the only included study to explicitly examine prevalence in a black population. Wong (1992) found levels of language delay for children from Hong Kong comparable to those in other studies.

Clinical judgement versus standardised procedure

We found no published study that attempted to integrate clinical judgement and standardised procedure with the effect that this might have on the estimation of prevalence. It is necessary, therefore, to interpret the results as being based on a notional psychometric convention. We would argue that this tends to lead to a relatively liberal cut-off which may overestimate the number of true clinical cases, particularly when seen in terms of a cut-off of more than 1 SD below the mean. However, it does not help us escape the essentially circular nature of statistically derived prevalence estimates. An alternative approach, and one which will be explored later in this report, is that prevalence should reflect the number of cases that the natural history would suggest are least likely to resolve spontaneously, and therefore most likely to be in need of intervention. It is doubtful if the estimation of prevalence will become more consistently accurate until the issue is addressed in this manner.

Prevalence and the screening process

The estimates of prevalence reviewed here suggest that speech and language delay in early childhood is a relatively common health problem. Whether such delay is, in fact, a sufficiently important health issue to merit screening is not addressed by the data. Even a moderate or mild delay may cause appreciable concern to parents and other carers. For a high-prevalence condition such as speech and language delay, the priority in terms of the screening process is to prevent the identification of false-positives while maximising the number of true-positives. This issue is the central focus of chapter 6.

The available evidence does not suggest that the prevalence of speech and language delay decreases over time. There are a number of reasons for this. It may be that children identified in the first few years remain delayed. For a screening programme this interpretation would suggest that such delays persist, and therefore identification and treatment of the children would be feasible. It may mean that some childrens' difficulties do resolve but that incidence effectively compensates for spontaneous recovery. Of particular interest is Silva's finding that, although the number of children with marked problems remains relatively constant given the cut-off adopted, the children themselves move in and out of the population of children with low language scores (Silva *et al,* 1983; for further discussion of this work see chapter 4). This would have implications for the identification of such children, in that some ages may represent a better opportunity for accurate diagnosis.

It may also be that the stability of prevalence estimates arises largely from study designs using standardised measures at a given cut-off level. The range of cut-offs adopted suggests that authors have attempted to define language delays operationally within a range of 1 SD, at –1 to –2 SD below the mean. Justification for the choice of cut-off is rarely provided, and it is difficult to say whether there has been any attempt to judge what is a manifest problem beyond reference to the psychometric property of the measure concerned. Tomblin (1997) is the only study to explicitly take his cut-off from clinical judgement. It is noteworthy that, with the exception of Paul and co-workers (1992), none have taken the view that only the most severe cases should be identified, suggesting a recognition that there is a need to cast the net sufficiently wide to draw in a group of children who are 'at risk' of persistent speech and language delays.

Summary

- The prevalence data originated from studies in which replicable criteria for case status was provided. In the majority of studies this involved the application of standardised assessment measures. All studies included in this review report results from normal rather than clinical populations.
- There is currently no common measure that makes it possible to synthesise these data except within those broad categories where commonality can be shown to exist. The range of prevalence estimates was 0.6% to 33.2%, the diversity reflecting whether specific or general speech and language delays were identified, the extent to which speech and language were combined, the nature of the population concerned, and the criteria used to define delay.
- Interpretation of the data is dogged by an uncertainty as to which children represent true cases in the sense that without intervention they would necessarily go on to experience problems associated with early speech and language delay in school and beyond. This issue is complicated further by social constructs of 'normality' and the difficulty in establishing a true gold-standard assessment with a theoretically motivated definition of case status.
- There are no data reviewed here addressing bilingual or (with the exception of three studies) ethnically diverse populations. It is also difficult, because of the nature of the data, to be clear about the relationship between social class and the prevalence of speech and language delay.
- The studies reviewed indicate a higher prevalence of speech and language delay in males than in females.
- There is little evidence to suggest declining prevalence across the 0–7 years age range but this is likely to be a function of the inclusion criteria adopted in this review. Identical cut-offs on standardised measures applied at different ages will necessarily result in the same proportion of the population being identified.
- There is no evidence to suggest that there is a real increase in cases in the period covered by the review (1967–97). This suggests that the estimation of prevalence and the demands made on services are not necessarily equivalent.
- It is not possible to give a definitive prevalence rate of the kind that would easily allow estimation of the adequacy of a screening procedure. This situation arises because of the different skill areas tapped in the prevalence studies, and due to the different, often arbitrary cut-offs used in determining a clinical case.

Chapter 4

The natural history of speech and language delay

Review questions

- What are the known outcomes for children with speech and language delays who do not receive systematic therapy services?
- What evidence is there that different subgroups of children with speech and language delays have different outcomes?
- Is there a difference in the pattern of outcomes reported from short-term natural history studies (e.g. 6 months) and those that span childhood?
- Is there a difference in the pattern of outcomes for children included as no-treatment controls and those in studies set up as natural history studies?
- Is there evidence from the natural history studies to suggest that children should be prioritised for intervention at different developmental stages and according to different presenting symptoms?

Inclusion/exclusion of literature

Relevance	Studies of children identified during the pre-school period as having a primary speech and language delay, who have not received any **specific** treatment or intervention for the condition.
Participants	Information about the number in sample.
Outcomes	Changes in speech/language measures (norm- or criterion-referenced) over time.
Design	(a) Prospective cohort studies of children who were identified as having primary speech and/or language delays but who did not receive speech and language therapy or other specialised input (b) Predictive validity trials of screening tests.

Large-scale prospective longitudinal studies potentially offer the best way of investigating natural history of outcomes. However, most of the prospective studies identified for this review were those in which at least a proportion of the subjects received speech and language therapy services.

A statement regarding the numbers in therapy, or the amount of therapy received, led to the study being excluded from this review. 'Therapy' was interpreted as specific advice or assessment from speech and language therapists. Where there was no such statement about therapy contact for the subjects, the study was included. In a minority of instances, a study explicitly stated there was no intervention for the subjects. General advice given routinely by a health professional was regarded as non-specific intervention, and such studies were included.

Consequently, some benchmark studies have been excluded (e.g. National Child Development Study of Sheridan and Peckham, 1975; Klackenburg 1980; The Portland Language Development Project, Paul *et al,* 1996). While their designs have many exemplary features and have achieved long-term follow-up for large samples, it has not proved possible to separate out possible treatment effects from their data.

Other criteria for inclusion ensured that there was a retest interval of at least 6 months, using norm-referenced or criterion-referenced language outcome measures. No minimum number of subjects was set.

Review of the data

There is extensive literature on the prognosis of children with speech and language delay. These follow-up studies highlight the adverse consequences for children whose language difficulties do not resolve by school entry. The studies fall into two broad categories: follow-up studies of those children with the most severe specific language delay who attended language units and other specialist provision (Petrie *et al,* 1975; Cooper *et al,* 1979; Urwin *et al,* 1988; Cook *et al,* 1989; Haynes and Naidoo, 1991) and community-based studies of language-impaired children (e.g. Bishop and Edmundson, 1987). The findings reveal that in addition to continuing problems in verbal lang-uage, reading, spelling and educational achievement more generally can also be affected (Aram and Nation, 1980; Stark *et al,* 1984; Bishop and Adams, 1990; Catts, 1991; Morris-Friehe

and Sanger, 1994; Tallal *et al,* 1997) together with behaviour and other aspects of psycho-social adjustment (Silva *et al,* 1987; Baker and Cantwell, 1987; Beitchman *et al,* 1989; Beitchman *et al,* 1996). While the problems are more marked for those children whose language difficulties are associated with low intellectual ability or which affect both receptive language and production, those with primary delay may also experience marked long-term difficulties which may persist to adulthood (Felsenfeld, 1992). However, except for the latter, none of these studies meet the review inclusion criteria.

The 12 studies that were reviewed were of three types. Nine natural history studies followed an identified but untreated group, with the study duration ranging from 8 months to 28 years. Sample size varied from four to 60 subjects across eight of the studies. The largest study followed an initial 1027 children from age 3 years through to 11 years (Silva *et al,* 1983). While most studies looked at the outcome of early language delay, only one natural history study focused on speech errors only (*Tables 5* and *6*).

The review also included three studies that examined the predictive outcome of an early screening procedure (Klee, 1997; Renfrew and Geary, 1973; Ward, 1992).

The median persistence figures for studies examining speech and language, speech only and language only are given in *Table 5* for all three study types together, and the individual studies are listed in *Table 6*. The median figures for studies examining expressive, receptive and expressive-receptive delays are given in *Table 7* and the corresponding individual studies are listed in *Table 8* (see also appendix 5 for more detail).

TABLE 5 *Median persistence reported for delay in speech and language, language only and speech only*

Type of delay [no. of studies]	Age range (years; months)	Median persistence (%)	Range of reported persistence (%)
Speech and language [1]	3;0–7;0	38	N/A
Language only [11]	0;10–7;0	66	0–100
Speech only [3]	4;10–33;0	50	22–100

N/A, not applicable

TABLE 6 *Persistence for individual studies of speech and/or language delay*

Study	Case studies/ original sample size	Persistence (%)	Age range (median [years;months])
Speech and language			
Fiedler *et al* (1971)*	46/138	38	3;0–7;0
Median		35	
Language only (expressive and receptive)			
Hall *et al* (1996)	5/9	100	4;7–7;0
Rescorla and Schwartz (1990)	25/25	54	2;2–3;0
Richman *et al* (1982)	22/705	65	2;0–3;0
Scarborough and Dobrich (1990)	4/16	0	2;6–5;6
Silva *et al* (1983)*	23/1027	78.2	3;0–7;0
Thal and Tobias (1992)	10/30	40	1;8–3;0
Klee *et al* (1997)	6/36	67	2;0–3;0
Ward (1992)	119/321	82	1;0–2;0
Ward (1992)	61/321	73	0;10–1;10
Ward (1992)	23/321	50	0;10–1;10
Median		66	
Speech only			
Bralley and Stoudt (1977)*	60/60	21.6	6;6–11;6
Felsenfield *et al* (1992)	24/52	50	4;10–33;0
Renfrew and Geary (1973)	150/150	54	5;0–5;6
Median		50	

* *Persistence for the longest period within the same study*

Outcomes for children not receiving systematic therapy services

Ward (1992) monitored infants from a mean age of 1 year through to 2 years of age. The sample was divided into three groups: expressive and receptive delay with listening difficulties; expressive and receptive delay without listening difficulties; and expressive delay alone. The outcomes were varied; while 82% of the first group continued to show language delay, the figures were 73% and 50% for the other two groups, respectively (*Table 6*). (Figures are for children with no associated developmental delay.) Of interest is the fact that the type of presenting delay changed in some cases from a combined expressive-receptive delay to one of expressive delay alone.

Studies following children with expressive-language delay from age 2 years, showed that 40–60% of children remained delayed in their expressive language achievement at age 3–4 years (Rescorla and Schwartz, 1990; Thal and Tobias, 1992). The study by Thal and Tobias included children of varying receptive language levels, while Rescorla and Schwartz specified receptive language within the normal range. Thal and Tobias noted a better outcome for those children with normal receptive skills and who used gesture to compensate for their lack of expressive output. Scarborough and Dobrich (1990) followed a small sample of four children from the age of 30 months through to 5 years. Although expressive language deficits resolved for all the subjects over this time, receptive language did not reach normal levels.

Studies of children followed from age 3 years, to age 4 or 7 years, also show the persistence of untreated language problems. Sixty-five per cent of expressive delays persisted to age 4 years

TABLE 7 *Median persistence for delay in expressive and receptive language, expressive language only, and receptive language only*

Type of delay [no. of studies]	Age range (years; months)	Median persistence (%)	Range of reported median persistence (%)
Expressive and receptive language [7]	0;10–7;0	75.6	26–100
Expressive language only [5]	0;10–7;0	40.0	0–54
Receptive language only [1]	3;0–7;0	8.7	N/A

N/A = not applicable

TABLE 8 *Persistence for individual studies of expressive and/or receptive language delay*

Study	Case studies/ original sample size	Persistence (%)	Age range (median [years;months])
Receptive and/or expressive language			
Richman *et al* (1982)	22/705	65	2;0–3;0
Ward (1992)	119/321	82	1;0–2;0
Ward (1992)	61/321	73	0;10–1;10
Hall *et al* (1993)	5/9	100	4;7–7;0
Klee *et al* (1997)	6/36	67	2;0–3;0
Silva *et al*[*] (1983)	23/1027	78.2	3;0–7;0
Median		75.6	
Expressive language			
Rescorla and Schwartz (1990)[†]	25/25	54	2;2–3;0
Thal and Tobias (1992)	10/30	40	1;10–3;0
Scarborough and Dobrich (1990)	4/16	0	2;6–5;6
Silva *et al* (1983)	21/1027	28.6	3;0–7;0
Ward (1992)	23/321	50	0;10–1;10
Median		40	
Receptive language			
Silva *et al*[*] (1983)	23/1027	8.7	3;0–7;0

[*] *Persistence for the longest of three periods within the same study*
[†] *Median between two expressive scales reported*

(Richman *et al,* 1982) and 38% to age 7 years (Fiedler *et al,* 1971). Within a full birth cohort, Silva and co-workers (1983) identified 8.4% with a language delay at age 7 years. Of particular interest is this latter study, which shows that those children presenting with a language delay represent a fluctuating group; some children fail at each of the three language assessment points, while some fail at only one or two assessment points. The more stable subgroup was that which included children with generalised language problems affecting both receptive and expressive skills.

The data from Hall (1996) are more limited. The study aimed to monitor language-disordered children from the age of 3;5–5;10 years through to age 7 years. (The focus of the study was the interaction of language and fluency skills.) Hall gave several language measures over time for individual subjects, but did not indicate explicitly which children continue to be designated as language delayed/disordered. The author indicated that all of the five children continued to have a language difficulty in at least one area of expressive or receptive skills.

Additional outcomes for areas other than spoken language are given by three studies (Scarborough and Dobrich, 1990; Richman *et al,* 1982; and Silva *et al,* 1987). These all point to reduced reading skills at age 7 or 8 years among those with earlier language delay (whether or not that oral language delay has resolved). In Silva's work, an early specific receptive delay was not associated with a later language delay, but 46% went on to have reduced reading or IQ measures at age 7 years. This group also went on to show the highest levels of behaviour problems at 9 and 11 years. Similarly, generalised early language delay was associated with low IQ, and poor reading and behaviour measures.

The studies looking at speech development included follow-up periods of 6 months, 5 years and 28 years, and all started with children who were about 5 years old. Over 6 months Renfrew and Geary (1973) showed that 54% of children with speech delay persisted, while Bralley and Stoudt (1977) reported nearly 22% of speech problems persisting over the course of 5 years. Felsenfeld (1992) followed some of the Templin longitudinal cohort (started in 1960) to track adulthood outcomes. In this study, children received no therapy until they were approximately 8 years old and there is no account in Felsenfeld's paper of any later therapy received. Felsenfeld found that 50% of children experienced residual speech problems, as assessed by sentence level tests of articulation. (Language measures in adulthood also showed skills deficits relative to controls, even though the children were originally identified as speech delayed.) More broadly, non-verbal reasoning and personality scores in adulthood were not found to be significantly different between the original speech-delayed group and the control group. It should be noted here that the controls were identified in adulthood and were not matched to those adults who had originally presented as speech delayed.

Outcomes for different subgroups of children with speech and language delay

Although several studies use 'expressive language delay' as an initial criterion, samples are varied in terms of receptive language skills and their non-verbal skills. Thus, Rescorla and Schwartz (1990) have a specific expressive delay (with normal range receptive language and normal range IQ), while receptive skills were not specified by Richman and co-workers (1982), Scarborough and co-workers (1990) or Thal and Tobias (1992). The latter two studies specified normal IQ, which was not the case for all of the samples of Richman and co-workers, and Ward (1992). Silva's data showed that early specific expressive delay is associated with later language delay, while children with specific receptive delay and generalised language delay had poorer outcomes for reading, IQ and behaviour (Silva *et al,* 1983).

Of issue in terms of generalisability is the nature of delays tracked by these studies. The severely or broadly language-delayed child is not well represented in the review data. This is a group that tends to have high levels of input from both educational and health services. Outcomes for these children are reported in the treatment follow-up literature and generally show a poor prognosis (see above).

Speech problems may be less persistent than language problems; Bralley and Stoudt (1977) showed that up to 78% of articulation errors resolve naturally. However, the long-term data from Felsenfeld (1992) suggest that underlying language difficulties may continue for children originally identified as having speech delay. Also, there is a body of evidence from follow-up studies of children treated for speech problems to suggest that literacy skills are at risk even after resolution of speech delay (Stackhouse, 1990).

Evidence for clinical prioritisation of intervention

Clinical prioritisation is a complex issue integrating local service needs with research findings. The data here suggest that children with expressive

and/or receptive language delays have more persistent and therefore arguably, more serious problems. This is not to deny the possible impact of more specific speech or expressive difficulties but suggests that the progression of such delays is particularly difficult to predict. However, it is possible to use the literature to pick out a number of factors that increase the risk of persistent delay. These include:

- **age**; children over 26 months have poorer expressive syntax outcome (Rescorla and Schwartz, 1990)
- **severity of delay**; the gap between estimated expressive language age (LA) and that expected for chronological age (CA) is correlated to expressive outcome a year later (Rescorla and Schwartz, 1990)
- **range of speech and language areas affected**; more generalised language delays affecting both expressive and receptive skills are more stable and cause more sequelae in the older school child (Silva *et al,* 1983). Also, when children with expressive delays have associated receptive delays, their outcome is likely to be poorer (Thal and Tobias, 1992)
- **general ability of the child**; children with low non-verbal skills and low verbal skills, show a high level of persistence of verbal delay (Richman *et al,* 1982)
- **associated factors;** these include neuro-developmental and medical factors, such as the incidence of pregnancy and labour complications, delayed first-year development, and hospital procedures during the first year, all of which have been found to be significantly higher in speech/language-delayed sample than in matched control children. Similarly sensorineural and conductive hearing loss should be considered risk factors (Fiedler *et al,* 1971).

The issue of protective factors, which might allow children to cope well with early speech and language delays such that they do not need therapeutic or educational support, has not been addressed in the data. It might be argued that isolated expressive difficulties especially at an early age might be considered a reduced risk. It is certainly true that many children with this type of difficulty improve spontaneously. However, it would not be appropriate to overstate the situation here. Many of these children do continue to have difficulties and to disregard their needs would be premature given the present data.

Summary

- The data set detailed above included only studies in which children received no specialist support for their speech and language delay. This data set is small even when predictive validity studies of over 6 months' duration are included. There is a much larger group of studies (not reviewed here) that provides information related to the follow-up of treated children.
- Spontaneous remission of speech and language delays is high in the pre-school period. Up to 60% of speech or language delays may resolve without treatment between the ages of 2 and 3 years. This figure is based upon studies which examined children with circumscribed expressive delays, often without evidence of serious difficulties in the first instance.
- The evidence suggests that, although prediction may improve with age, it remains somewhat imprecise and as such remains an issue of considerable significance to the identification process.
- The potential to predict outcomes increases if children experience expressive and receptive delays or more general developmental difficulties.
- Whether or not these oral language delays have resolved, multiple educational and social difficulties are noted for children who had earlier speech or language delays. Between 41% and 75% of early expressive language-delayed children showed reading problems at age 8 years. This finding is confirmed by the larger body of literature (not fully reviewed here) from follow-up studies of children who did receive intervention for speech and language delay.
- Individual outcome varies according to several identified risk factors, though these have yet to be successfully linked to a screening instrument.

Chapter 5

The effectiveness of intervention approaches for speech and language delay

Review questions

- What evidence is there that interventions can be shown to be effective when compared with untreated controls and other interventions?
- For which sub-groups of children (characterised by age and communication skills) has intervention been shown to be most effective?
- What evidence is there for the role played by associated difficulties (e.g. behaviour) in determining outcomes?
- Is there evidence that intervention for speech and language delay can be cost-effective?
- What components of the treatment process have an optimal effect?
- Do effect modifiers mitigate against drawing useful comparisons between studies?
- To what extent do the outcomes adopted reflect those recommended by the WHO (i.e. impairment, disability and handicap)?

This section of the review investigates the extent to which cases identified by screening for speech and language delay may be regarded as 'treatable'. Disorders of voice, stammering, cleft palate, sensory impairment (including hearing impairment), and neurological conditions which would be evident without screening were thus excluded.[1]

The compensatory education literature, notably studies from the US project Head Start and its successors, which attributed educational underachievement to adverse effects of social disadvantage upon language development, was also excluded as it did not meet the criterion of primary language delay. However, in view of the importance of educational considerations, it may be useful to note the key findings from such studies, particularly as many of the issues identified by these studies are relevant to our consideration of the effectiveness of intervention for primary language delay (e.g. the age at which treatment should commence, the intensity of intervention, the role of parents, and the cost-effectiveness of programmes).

Initial reports from Head Start projects indicated that, while vulnerable socially disadvantaged children made progress in response to intervention, these initial gains tended to 'wash-out' over time, leaving treatment groups no better off than controls (Cicirelli, 1969; Bronfenbrenner, 1974). However, the findings from long-term RCTs following participants through their school years and beyond, revealed the effectiveness of early intervention: high-quality, intensive pre-school education in a nursery setting allied to active parental involvement can result in long-lasting benefits in academic achievement and social adjustment, which persist to adulthood (Lazar and Darlington, 1982; Ramey and Landesman Ramey, 1992; Zigler and Muenchow, 1992). In addition, the Head Start literature provides evidence for the cost-effectiveness of early intervention. Schweinhart and Weikart (1993), for example, reported that prospective longitudinal studies extending over 25 years indicate that their programmes (the High/Scope Pre-school Curriculum Study Project and the Perry Pre-school Project) deliver four times the value of their initial costs as a result of long-term savings in education and social services budgets (i.e. children in the intervention groups were more likely to remain in mainstream education, graduate from high-school, enter employment and avoid delinquency).

Inclusion/exclusion of literature

Relevance	Studies of the effects of treatment/intervention upon primary speech or language delay in children up to 7 years of age.
Participants	Information about the number of participants in each group.
Outcomes	Comparison of pre- and post-intervention speech and language measures.
Design	(a) Experimental studies (RCTs) (b) Quasi-experimental studies (with non-random/pseudo-random control groups or non-equivalent control groups) including multiple time series studies with non-equivalent control groups (c) The following single-subject 'experimental' designs where there is no generalisation to untreated

control processes: withdrawal and reversal designs (ABAB, BAB, ABA), multiple baseline designs (across behaviours, settings or subjects or multiple probe variant) and alternating treatment designs, all with graphical displays or session-by-session data for individual subjects (baselines should have > 2 points with the exception of multiple baseline designs where **one** of the baselines may have 2 points).

Well-designed RCTs provide the strongest and most widely-recognised evidence for treatment efficacy (Crombie and Davies, 1996). Quasi-experimental designs, often a result of 'real-world' constraints upon the assignment of subjects to treatment or control conditions, provide a further source of evidence that can be cross-validated with the results from RCTs. However, in view of the effects of maturation noted in the previous chapter, the further requirement of an untreated control group was stipulated for both RCT and quasi-experimental designs in this study. (Note that these criteria excluded a number of well-known RCTs such as Barrett and co-workers (1992), which did not include an untreated control group, and Best and co-workers (1993), whose criterion for selection for intervention was broader than primary language delay).

Single-subject experimental designs have been widely used by practitioners and researchers in the area of language intervention since the 1980s, and provide evidence of a different character which can also be cross-validated with the findings from other designs. Three of these designs (withdrawal and reversal designs, multiple baseline designs, and alternating treatment designs) met the criteria for inclusion on the grounds that they specifically provide high levels of experimental control for the effects of maturation (Barlow and Herson, 1984; McReynolds and Thompson, 1986; Kearns, 1986; Connell and Thompson, 1986; Ingham, 1990; and also appendix 4 for further details). Full details of the basis for determining study validity and reliability may be found in appendix 8.

Framework for the analysis

RCT and quasi-experimental group designs

The results from the RCT and quasi-experimental intervention studies were synthesised by converting the outcomes into standardised effect sizes (Rosenthal, 1994) to permit comparison across studies (*Box 1*).

BOX 1 Standardised effect sizes

Standardised effect sizes provide a means of converting the outcome measures from studies into a common form which can be combined across different studies. In the case of the RCT and quasi-experimental studies here, the effect size used was *d*, the difference between the post-test means of treatment and non-treatment control groups divided by the pooled SD for each study, corrected for population effect size bias, which particularly effects small samples (Rosenthal, 1994).

Effect sizes with positive signs indicate that subjects in treatment groups achieved higher post-intervention scores that those in non-treatment control groups, that is, that there was a positive treatment effect. Conversely, effect sizes of zero, or with negative signs indicate studies in which subjects in treatment groups failed to make greater progress than those in control groups.

While effect sizes can also be computed from 'gain' or difference scores (i.e. post-test–pre-test) there are problems with this approach. As Sheehan and Gallagher (1983) note, gain scores are less reliable than actual performance scores (see Lord, 1956 and Cronbach and Furby, 1970 for a rationale in terms of measurement theory) and in addition, commonly have a negative correlation with pre-test scores (Cohen and Cohen, 1983). Pre-test scores and gains were significantly correlated in 11 of the RCT/quasi-experimental studies in the present data set. Pre-test–gain score correlations were also evident in five studies as a result of the use of analysis of covariance, with pre-test scores as covariates (Almost and Rosenbaum, 1998; Fey *et al,* 1993; Gibbard, 1994 [Studies 1 and 2]; Fey *et al,* 1994) and inspection of the raw data revealed significant correlations in a further three studies (Wilcox and Leonard, 1978; Zwitman and Sonderman, 1979; McDade and McCartan, unpublished). A further three studies did not report sufficient data for a gain score to be calculated (Ward, 1994 [Groups 1 and 2]; Methany and Panagos, 1978). Effect sizes were thus calculated using post-test scores for treatment and non-treatment groups. There was one exception to this procedure; Shelton and co-workers (1978) in both Study 1 and Study 2, reported their negative findings in the form of gain scores only. To minimise the possible effects of bias that might result from discarding known negative effects, effect sizes calculated from *t*-test values were included in the present analysis. Two studies from the data set, Conant *et al* (1984)

and Reid *et al* (1996) were excluded from further analysis as it was not possible to derive their effect sizes.[2]

Single-subject designs

The multiple baseline across subjects design is one of three single-subject designs included in the present review and is illustrated in *Figure 2*. The first element is the baseline, which extends over a period of sessions, followed by the treatment, which is introduced once a stable baseline has been established. Intervention may be followed by a period of maintenance or follow-up during which treatment is withdrawn. In the example here, the design is replicated across more than one subject with treatment introduced to each subject sequentially. The treatment is assumed to affect only the behaviour targetted by the intervention, and other baseline or control behaviours should hence remain unchanged. The stability of the baselines for Subjects 2 and 3 (after the onset of treatment for Subject 1) and their subsequent changes in response to the introduction of treatment provide not only evidence that the observed changes in performance are the result of the treatment, but also experimental control for the effects of maturation. Single-subject experimental designs provide data regarding the generalisation of treatment to other settings, and the maintenance of any gains after the withdrawal of treatment, two key issues underpinning the effectiveness of intervention. However, although the use of individual subjects as their own controls can reduce considerably the variability in a study, there are problems in generalising findings because of the small samples used.

The data from the single-subject experimental designs were analysed using the PND statistic (Scruggs *et al*, 1988) (*Box 2*).[3]

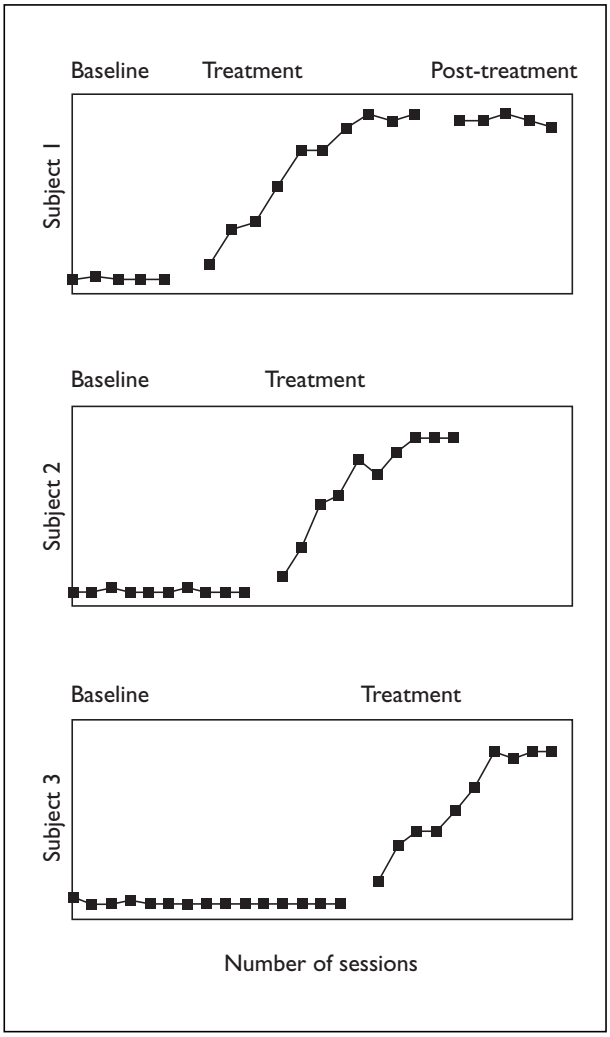


FIGURE 2 *Illustration of a single-subject experimental design (multiple baseline across subjects)*

BOX 2 Percentage of non-overlapping data

The synthesis of data from single-subject designs requires a different approach to the standardised effect statistic. The metric used was the PND between baseline and post-baseline phases. This was used to carry out an exploratory synthesis of the results from the single-subject studies. The PND provided ordinal measures of the extent to which post-baseline points show an increase following the introduction of treatment. PND scores can be computed for generalisation and maintenance/follow-up phases as well as for treatment and can be analysed using non-parametric tests such as the Kruskal Wallis and Mann-Whitney tests. The percentage of post-baseline points on the graphical display of baseline and treatment phases which are above the highest baseline point is calculated, and the result is the PND.

Review of the data

Comment on the quality of studies

The literature searches identified a data set of 48 studies of the effectiveness[4] of intervention which met the inclusion/exclusion criteria. Ten of these used RCT designs, 12 quasi-experimental designs[5] (all with the non-random allocation of individuals to treatment or non-treatment groups) and 26 single-subject designs. Of the single-subject designs, 22 had sample sizes of two or more. Details of the included studies are summarised in appendix 5. *Table 9* summarises the key qualities of the RCT and quasi-experimental studies in the data set in terms of factors which are known to affect study validity (Cook and Campbell, 1979). (A full breakdown of these data is presented in appendix 9.)

TABLE 9 *Summary of quality factors for RCT and quasi-experimental designs*

Study design	No. of studies	Subject recruitment	Subject allocation to groups	Blinding of assessors	Comparability of pre-test scores
RCT	10	Selected: 60% Non-random: 30% Selected and non-random: 10%	All random	Yes: 40% Not stated: 60%	Yes: 70% No: 20% Not stated: 10%
Quasi-experimental	12	Selected: 100%	All non-random	Yes: 8% Not stated: 92%	Yes: 76% No: 16% Not stated: 8%

TABLE 10 *Summary of quality factors for single-subject designs*

Study design	No. of studies	Use of untreated control processes	Baseline stability*	Generalisation outcomes	More than one subject
Single subject (validity score at least 10)	17	3 studies (18%)	13 studies (76%)	14 studies (82%)	14 studies (82%)
Single subject (validity score below 10)	9	1 study (11%)	3 studies (33%)	7 studies (78%)	8 studies (89%)

* *Percentage of studies with at least half of their baselines stable*

Three of the RCTs dealt with more than one area of language, thus providing comparative data across language domains. While none of the RCTs used random sampling from a population of children with language delay, the children were randomly allocated to treatment and control groups. Assessors were 'blind' to case status in only four studies, but pre-test scores were comparable across groups in all but three studies. Reliability and validity scores were high (i.e. > 75% of the maximum possible) in seven out of the ten studies. In the quasi-experimental studies five involved more than one area of language. Allocation of children to treatment and control groups in the quasi-experimental studies was non-random, which adversely affects the internal validity of the studies. However, pre-test scores were comparable across conditions in eight of the 12 studies. Reliability and validity scores were lower for the quasi-experimental designs than for RCTs, with seven of the 12 studies falling below a level of 75% of the maximum possible score for reliability, and nine falling below a similar level for validity.

Table 10 provides information about the study quality of the single-subject designs. For ease of presentation these are summarised as studies which have an above-average validity rating of 10 or above (out of a maximum of 18, see appendix 8) and those with a below-average rating.[6] The key issues here are the presence or absence of untreated control processes designed to remain invariant across treatment (hence providing additional experimental control), the length and stability of the pre-intervention baseline, the number of treatment sessions, the presence or absence of opportunities for generalising the outcomes of intervention to unfamiliar settings, behaviours or materials, and finally replication across subjects.

The single-subject designs reviewed were more focused than the group designs, with all but two studies dealing with only one area of language. Multiple baseline designs (across behaviours in 11 studies and across subjects in eight) were the most widely-used designs. Six studies used alternating treatment designs (in three cases linked to multiple baselines). One study used the withdrawal design (again, linked to a multiple baseline) which provides a measure of experimental control over possible placebo effects. Only four studies overall used untreated control processes.

Baselines in the higher quality studies varied in length from 3–25 sessions. Ideally, a baseline should remain parallel to the abscissa with only minimum variability but there is no generally accepted criterion for stability of baseline.

Sidman (1960) has proposed no more than a 5% range of variability. However, others have argued for variability of less than 10% across at least three data points (Bain and Dollaghan, 1991; Powell *et al,* 1991). A stable baseline was defined operationally as one in which all scores remain within a range of ± 10% of the mean baseline score. Using this criterion, only five of the single-subject studies with higher validity had stable baselines for all of their dependent measures. The length of the treatment phase varied considerably across the high-quality studies (from 4–70 sessions) but no fewer than 14 of these studies provided measures of generalisation and only three had one subject.

Only one of the studies which achieved lower validity rating scores used an untreated control process, and overall the range of length of baselines (3–10 sessions) was smaller than in the higher quality studies. Baseline stability was also poorer and there were fewer treatment sessions (from 6–20). However, all but two have generalisation measures and eight had more than one subject.

These three classes of design, RCT, quasi-experimental and single-subject formed a hierarchy of evidence based upon study validity. The quasi-experimental and single-subject designs, although accorded lower quality ratings, provide supporting evidence for the small set of RCT designs.

The effectiveness of interventions compared with untreated controls

Reviews of the literature (e.g. Guralnick, 1988; Olswang and Bain, 1991; Law, 1997; McLean and Woods-Cripe, 1997; Zwart, unpublished) identify a number of key variables that should be considered when the effects of intervention are being evaluated. These include child variables (age, gender, social class, nature of presenting difficulties) and programme variables (area of language treated, direct or indirect treatment by the clinician, model of intervention, intensity and duration of treatment), as well as the role of parents or care-givers. Appendix 10 provides a full breakdown of details from the studies in terms of these parameters. *Table 11* summarises the outcomes from studies which used RCT or quasi-experimental designs. (Descriptions of the different intervention approaches as classified here can be found in appendix 2.) Note that intervention models are classified here in terms of their presentation to the **child** and, unless otherwise stated, the frequency of treatment refers to the clinician's contribution. Thus, an approach which uses significant others in the child's environment (e.g. parents, teachers) would be classed as indirect because the speech and language therapist is working via an intermediary.

The results from RCT studies

The RCT studies reviewed provide examples of:

- intervention in the areas of articulation/phonology, expressive language, receptive language, auditory discrimination/listening skills and parent–child interaction
- early versus later intervention (with four studies involving children whose average age was less than 3 years)
- intensive versus less intensive therapy (with three studies offering three or more sessions per week)
- didactic treatment (four studies) and hybrid treatment approaches (six studies)
- direct and indirect treatment (with six of the studies involving parents, of which two also provide a direct comparison of the outcomes from clinician- and parent-administered intervention).

The overall results reveal statistically significant treatment outcomes ($p < 0.05$) for nine of the ten studies across the three areas of language, despite the relatively small numbers of subjects involved.[7] The two direct comparisons found parent-administered treatment to be as effective as direct treatment by the clinician (Fey *et al,*

TABLE 11 *Summary of treatment outcomes for RCT and quasi-experimental designs*

Study design	No. of studies	Areas of language intervention*	Mean age of treated children (months) [range]	Proportion of studies with significant results
RCT	10	1, 2, 3, 4, 6	42 [23–70]	90%
Quasi-experimental	12	1, 2, 3, 4, 5, 6	39 [8–98]†	83%

* *1 = phonology/articulation; 2 = expressive language (syntax and/or vocabulary); 3 = receptive language (comprehension and/or vocabulary and/or auditory association); 4 = auditory discrimination/listening skills/phoneme awareness; 5 = pragmatics; 6 = parent–child interaction*

† *Three of Wilcox and Leonard's (1978) sample (n = 24) were aged between 7;0 and 8;2. It was decided to include them rather than discard the study*

1993; Gibbard, 1994), and both of the studies evaluating outcomes from the Hanen parent training programme yielded significant results (Girolametto *et al,* 1995; Girolametto *et al,* 1996). In addition, there was evidence of a two-way transfer of training in syntax to phonology and vice versa (Methany and Panagos, 1978), though one short-term intensive parent-administered programme failed to show any effects of generalisation from auditory training to improved articulation (Shelton *et al,* 1978, Study 1). However, only one of the studies, (Lancaster, 1991), provided any information about normalisation of subjects following treatment: in total six of the children in the treatment groups had post-test scores that were within the average range for their CA.

The results from quasi-experimental designs

The studies using quasi-experimental designs provide examples of:

- intervention in the areas of articulation/phonology, expressive language, receptive language, auditory discrimination/phoneme awareness, pragmatics and parent–child interaction
- early versus later intervention (four studies involving children whose average age was less than 3 years)
- intensive versus less intensive therapy (though only one study offered more than three sessions per week)
- didactic treatment (four studies), naturalistic treatment (one study) and hybrid treatment approaches (seven studies)
- direct and indirect treatment (with four of the studies involving parents, two also providing a direct comparison of the outcomes from clinician- and parent-administered intervention).

As in the case of the RCTs, the results here reveal statistically significant treatment outcomes ($p < 0.05$) for ten of the 12 studies, with substantial treatment effects in a further two which did not report any statistical analysis.[8] Significant outcomes were observed in all six of the areas of language in which intervention took place. As before, the only non-significant finding was that of Shelton and co-workers (1978, Study 2), where auditory training did not generalise to improvements in the target sound production of children with problems in articulation and phonology. Parent-administered treatment was effective in three out of four cases, and where a direct comparison was possible, resulted in outcomes that did not differ from those of direct treatment.

These studies extend the range of findings from the RCTs in the following ways.

- More areas of language are represented (e.g. receptive language and pragmatics).
- A broader age-range (including examples of highly effective early intervention in the child's first year) is included.
- An example of a naturalistic intervention approach is included.
- Evidence of normalisation following treatment is provided; McDade and McCartan (unpublished), Ward (1994), Warrick and co-workers (1993), and Whitehurst and co-workers (1991), report that significantly more of the treated children than controls scored within the average range at post-test.
- More follow-up data are reported than in the case of the RCTs: some studies report that gains are maintained for over a year after treatment (Ward, 1994; Warrick *et al,* 1993) while one reports the 'wash-out' of early gains over time (Whitehurst *et al,* 1991).[9]

Notwithstanding the significant results reported, sample sizes across studies were again modest, with five studies averaging less than ten children per group. In addition, while 14 studies (seven RCTs and seven quasi-experimental designs) reported the use of normative tests, the authors in all but three cases appear to have carried out their analyses on the raw scores rather than on standard scores, age-equivalent or percentile scores, thus losing a measure of protection against the effects of maturation during the test–post-test interval.[10]

The results from single-subject designs

The findings from the single-subject designs provide further confirmation of the effects of intervention in the areas of articulation/phonology, expressive language, receptive language, and are based on outcome measures which are closely linked to the process of treatment. There is evidence of the effectiveness of didactic, naturalistic and hybrid intervention approaches across language domains and also of the effectiveness of indirect treatment, administered by parents (in three studies) and by teachers (one study). The studies also provide evidence of the generalisation of treatment outcomes and of the maintenance of gains following the withdrawal of intervention.

Synthesis of treatment outcomes

The analysis of treatment outcomes was further explored using effect sizes (RCT and quasi-experimental data) and PND measurements (experimental single-subject data). *Table 12*

TABLE 12 *Summary of effect sizes of included studies by study design*

Design	Total no. of subjects	Mean CA (months) [range]	Treatment characteristics	No. of effect sizes
RCT	250	42 [23–70]	**Direct treatment by clinician** median 9 hours per child, in 21 30-min sessions over 4 months	10
			Indirect treatment* median 17 hours of clinician time per child in ten 90-min sessions (usually group) over 5 months	50
Quasi-experimental	368	39 [8–98]†	**Direct treatment by clinician** median 14 hours in 21 40-min sessions over 5 months	23
			Indirect treatment median 19.5 hours of clinician time in 11 90-min sessions (usually group) over 4–5 months	26
Single-subject	73	52 [23–83]	**Direct treatment by clinician** median 12 hours in 18 45-min sessions over 3 months	153‡
			Indirect treatment median 12 hours of clinician time in 14 50-min sessions (usually group) over 2–3 months	16

* *Indirect treatment by clinicians, for example, parent- or teacher-administered*

† *Three of Wilcox and Leonard's (1978) sample (n = 24) were aged between 7 years and 8;2 years. It was decided to include them rather than discard the study*

‡ *Strictly, PNDs provide a measure of effect rather than 'effect size'* per se

TABLE 13 *Summary of effect sizes by language area and by direct/indirect treatment: study level analysis*

Language area	Norm-referenced measures (95% CI)				Criterion-referenced measures (95% CI)			
	n*	Direct	n*	Indirect	n*	Direct	n*	Indirect
Articulation/phonology	2	+1.11† (+0.46/+1.77)	2	+0.20 (–0.44/+0.83)	3	+0.94† (+0.37/+1.52)	4	–0.02 (–0.52/+0.47)
Expressive	5	+0.65† (+0.23/+1.10)	9	+1.08† (+0.83/+1.34)	4	+1.11† (+0.58/+1.63)	5	+1.16† (+0.75/+1.56)
Receptive	2	–0.02 (–0.66/+0.63)	5	+1.43† (+1.09/+1.77)	N/A	N/A	N/A	N/A

* *The number of studies which contributed an effect size*

† *Indicates statistically significant results (p < 0.05)*

provides a summary of the number of effect sizes for each of the three study designs.

A separate meta-analysis (Rosenthal, 1994) was carried out for each component area of language for which data were available (i.e. articulation/phonology; expressive language, including syntax, semantics and vocabulary; receptive language, including comprehension and vocabulary; and auditory discrimination/listening skills/phoneme awareness). This circumvents many of the problems of combining effect sizes across multiple measures of different dependent variables and ensures independence of measures (Lipsey, 1994; Rosenthal, 1984). Separate analyses were carried out for norm-referenced and criterion-referenced measures, as the latter generate higher effect sizes (Nye *et al*, 1987). Different treatments were analysed separately ('best' treatment combination versus 'worst' treatment combinations) but similar multiple criterion-referenced measures from one treatment within a given domain were averaged to yield a single combined measure (Rosenthal, 1994). Effect sizes from the RCT/quasi-experimental designs were analysed following the procedures recommended by Hedges and Olkin (1985).[11]

Full details of the meta-analysis are provided in appendix 11 but the outcomes are summarised in *Table 13*. Due to the presence of confounding

factors (i.e. variables such as age, gender, language area, intervention approach and study design which vary with each other so that the results ostensibly due to one variable could be due to another[12]) the only comparison possible across studies was of direct versus indirect treatment. The results reveal the effectiveness of direct and indirect treatment approaches for expressive language and receptive language across both norm-referenced and criterion-referenced measures. However, only direct treatment was effective in the case of articulation/phonology, though the small number of studies in this area and the use of non-standard treatment approaches in the indirect treatment condition should be noted.

While these findings provide overall support for the effectiveness of intervention, the relatively small number of studies should be noted, particularly in the case of direct treatment for articulation/phonology and receptive language problems. However, note that an effect size of +1.00 corresponds to a level of progress equivalent to that from the 5th to the 25th percentile on a norm-referenced test, a considerable degree of normalisation.

Effect sizes from studies with RCT/quasi-experimental designs were also averaged across language areas to allow a comparison between the data from children with primary speech and language delay and those from meta-analyses of studies involving children with secondary delay (*Table 14*).

Data comprising all of the effect sizes across studies and effect sizes from higher quality studies only (with validity scores of > 10) are presented separately. The results from the controlled studies here are comparable to those reported by Nye and co-workers (1987) which included a wider range of study design and study quality. Nye found an average effect size of +1.42 from 23 effect sizes for outcomes in syntax (which corresponds closely to the figure above for expressive language), and an average effect size of +0.65 for comprehension, from 13 effect sizes. (Shonkoff and Hauser-Cram (1987) also reported an average effect size of +1.17 for language outcomes from 31 studies of intervention with disabled children.[13])

These findings are somewhat higher than the estimates reported by Casto and Mastropieri (1986) (mean effect size +0.67) and by Arnold and co-workers (1986) (mean effect size +0.59). However, these latter results are difficult to interpret because the effect sizes are based on a wide range of measures, including IQ, and factors such as severity of the handicap are confounded with the type of intervention programme and other variables, such as age.

PND statistics were calculated from the studies employing experimental single-subject designs for treatment, generalisation and maintenance/follow-up phases for 73 subjects. Only one PND was calculated for each effect (treatment, generalisation or maintenance) for each subject and each type of effect was analysed separately, thus avoiding problems of multiple measures.[14] A summary of outcome effects for articulation/phonology and expressive language is shown in *Table 15*.

Generalisation of treatment and the maintenance of gains after the end of intervention are important indicators of effective outcomes. The results from the single-subject designs reveal interesting findings regarding the relative effectiveness of intervention approaches and of different types of generalisation. Hybrid intervention approaches (five subjects), for example, yielded significantly higher levels

TABLE 14 *Summary of effect sizes by language area: effect size level analysis*

Area of language	No. of effect sizes	Average effect size	95% CI
Articulation/phonology (all studies)	29	+0.35*	+0.10/+0.60
Study validity score > 10	15	+0.85*	+0.60/+1.10
Expressive language (all studies)	57	+1.07*	+0.85/+1.29
Study validity score > 10	27	+1.20*	+0.98/+1.42
Receptive language (all studies)	7	+1.09*	+0.44/+1.74
Study validity score > 10	N/A	N/A	N/A
Auditory discrimination (all studies)	14	+0.23	–0.10/+0.56
Study validity score > 10	N/A	N/A	N/A

* *Indicates statistically significant results (p < 0.05)*
N/A = not available

TABLE 15 *Summary of PND outcomes for single-subject designs*

Area of language	Treatment*		Generalisation		Follow-up	
	No. of subjects	Median PND	No. of subjects	Median PND	No. of subjects	Median PND
Articulation/ phonology	18	87.72	31	75.00	13	100
Expressive language	24	81.74	24	87.50	8	96.43

Note: Articulation/phonology and expressive language accounted for all but four of the observed effects in these studies

* *Outcomes here are for the most effective treatment per study*

of maintenance than didactic approaches (three studies) in the case of expressive language ($Z = 1.92$, $p < 0.055$), and naturalistic approaches resulted in marginally more significant levels of generalisation than didactic in the case of articulation/phonology ($Z = 1.79$, $p < 0.08$).

With regard to transferability of training, generalisation across setting (e.g. from clinic to home) or across different behaviours led to better outcomes than generalisation to different, untrained stimuli or untrained stimuli in a different setting ($Z = 2.62$, $p < 0.01$). Similar findings have been reported by Scruggs and co-workers (1988). Higher quality studies also tended to generate better treatment outcomes (Spearman's rho = +0.45, 42 df, $p < 0.002$).

These results require cautious interpretation due to the nature of the PND data and the small numbers involved in some of the comparisons, but they provide confirmatory support for the effectiveness of intervention overall. The results also suggest specific research questions that could be explored using controlled studies, for example, "Are naturalistic/ hybrid intervention approaches more effective in the long term than didactic approaches?" "Is specific training required to maximise generalisation of the effects of treatment?".

In summary, the results from the meta-analysis carried out on post-test scores from the reviewed studies confirm that intervention can be effective for problems in articulation/phonology, expressive language and receptive language. There were too few studies to synthesise the findings from treatment of problems in auditory discrimination and phoneme awareness, but a training programme in the latter area generalised to progress in reading with gains that were evident a year later (Warrick *et al,* 1993). The strongest evidence for effectiveness comes from the intervention studies in expressive language where the higher number of studies involved in the analysis added to the reliability and generality of the findings.

Where is intervention most effective?

Ten of the studies with RCT/quasi-experimental designs involved children with specific expressive language delays, seven studies involved children with specific articulation/phonological problems and only five involved children with mixed receptive/ expressive delay. While the original studies report significant gains in these areas it was not possible to compare the outcomes directly across sub-groups because of the small number of subjects and the presence of confounding variables.

Behaviour and outcome

Behavioural difficulties are under-specified in the present sample. Only one study (Girolametto *et al,* 1995) reported a reduction in acting-out behaviour in a small experimental group of eight children following language intervention. The only other reference to behaviour was in the Hemmeter and Kaiser (1994) study where one child had a history of behavioural problems.

The cost-effectiveness of intervention for speech and language delay

The finding that indirect treatment can be as effective as direct treatment for problems in expressive language and receptive language has implications for cost-effectiveness. *Table 12* provided an indication of the potential savings in clinician time that might result from indirect intervention. However, no information was available from included studies regarding the scale of possible savings in special education and other support services as a result of early intervention with children with primary language delay.

Two studies not reviewed have provided detailed costings for direct versus indirect treatment of language handicaps in the USA, including costings of parental time and transport (Barnett *et al,* 1988; Eiserman *et al,* 1990). Based on their 1987–88 data, the latter study showed a cost per service hour of US$64 for clinic-based intervention for children aged 3–5 years with 'moderate' speech disorders

(i.e. scores below the 20th percentile on a standardised test of articulation) compared with US$31 for home parent training. Barnett reported similar findings based on data for a sample of children aged 3–5 years with speech and language delay/disorder. There is a need for UK-based studies to investigate cost-effectiveness issues of direct versus indirect treatment.

The effectiveness of treatment components

There is evidence from individual studies of the effectiveness of a range of didactic, naturalistic and hybrid treatment approaches (see appendix 10). However, the data from the RCT and quasi-experimental designs do not permit direct comparisons of components of the treatment process due to the small size of the data-set and the presence of confounding variables. There is tentative evidence from the single-subject design studies regarding the relative effectiveness of naturalistic/hybrid intervention approaches compared with didactic approaches, but this requires further investigation from controlled studies.

The influence of effect modifiers on study comparisons

The small number of controlled studies and small samples reduced significantly the number of comparisons possible between studies, but in any event, many effect modifiers (e.g. treatment characteristics, design characteristics and child characteristics) were confounded. In particular, it was not possible to determine directly the effects of age upon treatment outcomes, which has important implications for the timing of intervention.[15]

Outcomes and the WHO

None of the studies reviewed to date have outcomes that reflect those recommended by the WHO (i.e. impairment, disability and handicap).

Data representativeness

Two checks were carried out on the representativeness of the data. First, 'funnel plots' (Light and Pillemar, 1984) comparing effect sizes by sample size (*Figure 3*) and effect sizes by study

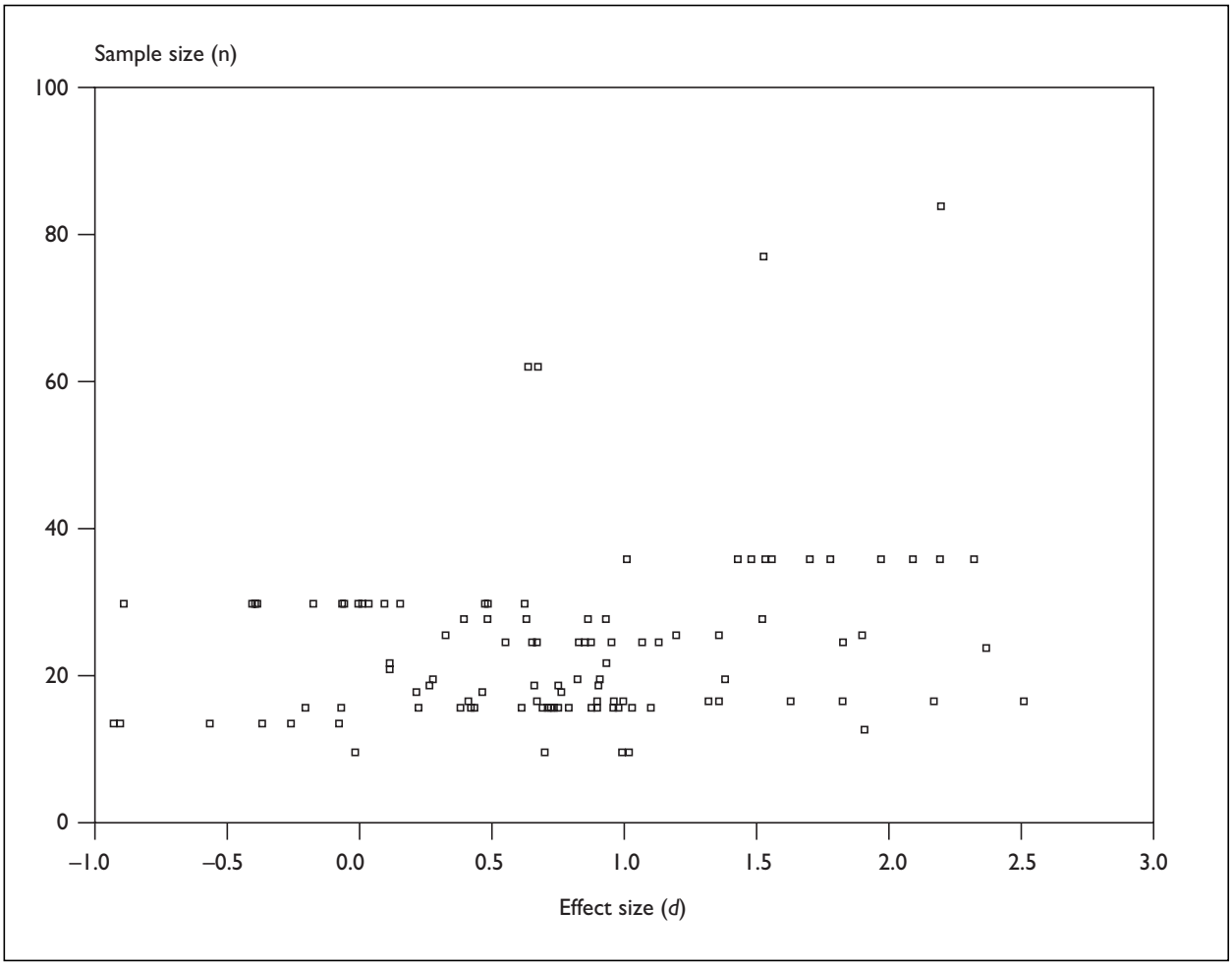


FIGURE 3 *Funnel plot of 109 effect sizes from RCT/quasi-experimental designs by sample size*

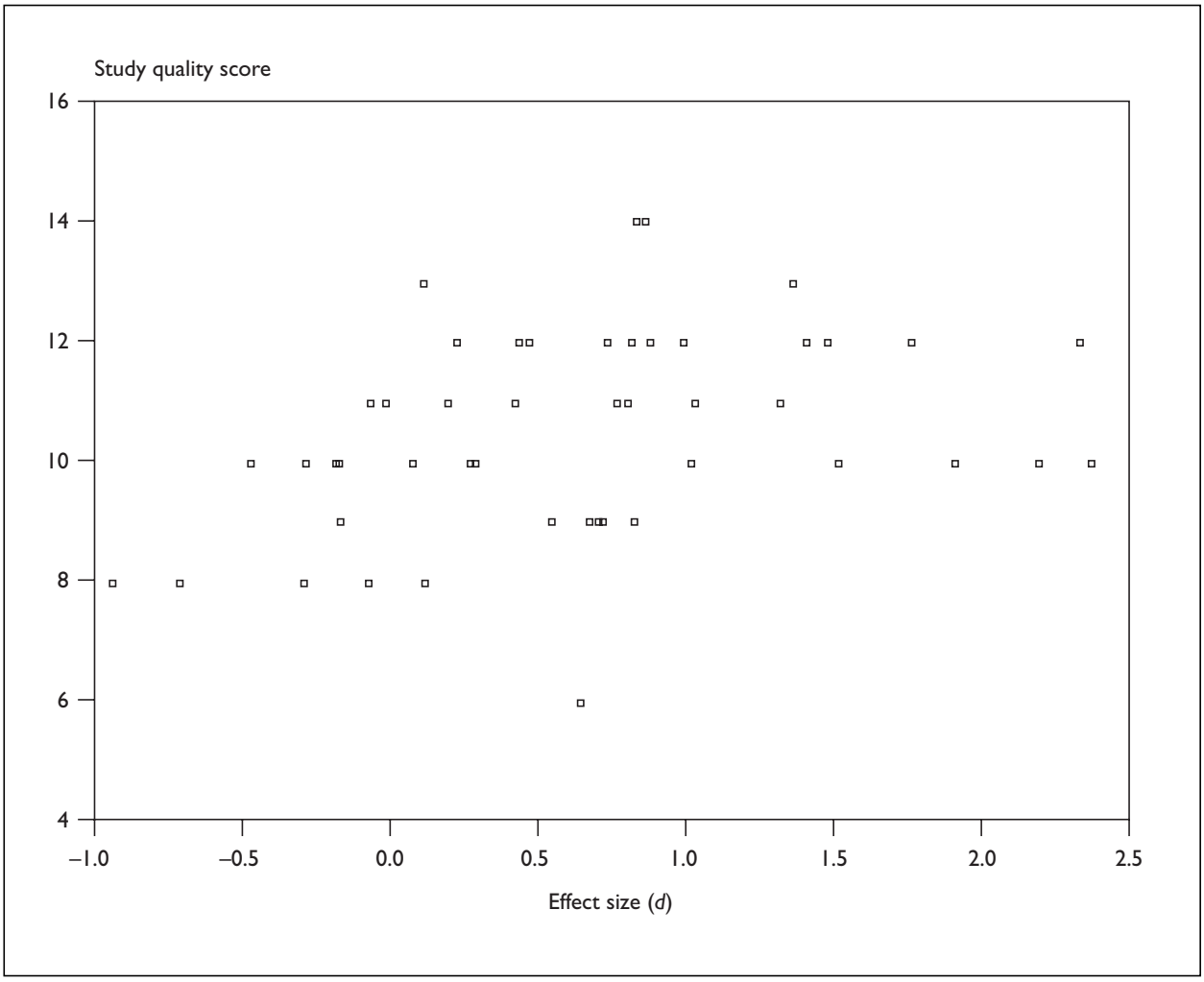


FIGURE 4 *Funnel plot of 109 effect sizes from RCT/quasi-experimental designs by study quality (i.e. validity)*

quality (*Figure 4*) are shown above. These plots are used as a visual diagnostic test to determine whether there are any 'gaps' in the literature that could be the result of publication bias (e.g. the extent to which only significant studies have been reported). There is evidence of some skewing due to the four effect sizes from the studies with comparatively large samples but no indication of marked 'gaps' in the data set that could be the result of publication bias.

A second check was carried out on excluded intervention studies to explore to what extent these studies also showed positive intervention effects. Of the 80 excluded studies (with 82 sets of data), 75% of the RCTs/quasi-experimental studies and 85% of the single-subject experimental studies showed positive effects, thus further reducing the likelihood that the intervention effects reported here are the result of Type I statistical error (i.e. concluding incorrectly that the children receiving intervention benefited).

Summary

- The findings from the reviewed studies confirm the effectiveness of intervention for primary speech and language delay in the areas of articulation/phonology, expressive language, receptive language and in phoneme awareness. The effect sizes from the studies overall indicate progress of the order of 1 SD corresponding to progress from the 5th to the 25th percentile on a standardised test. The single-subject studies also provide evidence of generalisation of treatment effects.
- There is evidence for the effectiveness of didactic, naturalistic and hybrid intervention approaches. However, direct comparisons of the effects of intervention in different areas of language, of different types of treatment, of intensity of treatment, and of the effects of treatment across different age-groups were not possible due to the small number of studies in the data set and the number of confounding

variables. It is not possible, therefore, to identify characteristics of treatment (e.g. timing, intensity, duration, setting, approach) that are optimal. Further studies are needed of the impact upon treatment effectiveness of age and types of treatment. However, caution is necessary; first, only eight of the 48 studies included here were carried out in the UK, and second, the extent to which the results from the RCT and quasi-experimental studies reflect everyday clinical practice is uncertain.

- The evidence from the meta-analysis reveals that indirect, parent-administered treatment in the areas of expressive language and receptive language is at least as effective as direct, clinician-administered treatment. In the case of articulation/phonology, however, direct treatment was found to be more effective, though non-standard treatments were used in the indirect condition and the comparison involved only seven studies. The extent to which the indirect treatment effects reported here reflect the levels of parental cooperation and compliance observed in everyday practice is also uncertain.
- Most of the reviewed studies dealt with problems in articulation/phonology and expressive language and there were fewer studies of delay in receptive language (two RCT studies and a further five quasi-experimental studies), only two of which involved direct treatment. This is problematic, given the association between difficulties in receptive language and persistence of language delay (Whitehurst and Fischel, 1994). More well-controlled research into the effectiveness of intervention for problems in receptive language is required (one such study will be reported in 1998/99 by Law, Kot and Barnett).
- The studies here do not provide any information that would help to distinguish 'late talkers' who are likely to catch-up, from those children with primary delay who are likely to have persistent language delay and experience long-term problems in educational achievement and social adjustment.
- Few studies examine the long-term effects of early intervention in the area of primary language delay. Whitehurst and co-workers (1991) report that the gains 'wash-out' over time, but differential attrition rates in their study pose problems for the interpretation of their findings. Controlled studies of the long-term effects of the early treatment of primary language delay with a focus upon subsequent school achievement are required. The sample identified by Ward (1994) before the children's first birthday and treated in the second year of life is due to be reassessed in school in the near future. This data should contribute considerably to the discussion of the outcomes for treated groups in the longer term.
- There is little use of normalisation data for post-test scores and no use of WHO outcomes. Future use of such data would greatly strengthen the body of evidence.
- Small sample size is a recurrent theme in this chapter. It is imperative that future studies address the issue of statistical power (Cohen, 1992). A large-scale study provides the most convincing evidence of overall effectiveness. A large RCT is currently underway in the UK (Roulstone and Glogowska) and will be reported in 1998/99.

Notes:

[1] For recent reviews of research into intervention for disorders of voice see Enderby and Emerson (1995, ch 9) and Sloane (1995); for stammering and stuttering see Conture (1996), and Lincoln and Onslow (1997); for cleft palate see Enderby and Emerson (1995, ch 4), and Russell (1995); for studies of the outcomes of hearing impairment see Mogford (1988), and Roberts and Scheule (1990); and for neurological conditions see Hall (1995), and Enderby and Emerson (1995, ch 5).

[2] Conant *et al* used a multivariate analysis of variance (ANOVA) (see Kraemer and Andrews, 1982), and Reid *et al* did not carry out a direct test of the difference between their experimental and control groups. We were unable to obtain means and SDs, which would have allowed us to calculate effect sizes.

[3] The possibility of calculating tests of trend for the single-subject data was considered (Olswang, 1990) but this approach is also not without its problems (Kratchowill, 1978; Scruggs *et al,* 1987). See White (1987) and Salzberg *et al* (1987) for a discussion of the PND.

[4] Effectiveness is used here to refer to whether a treatment or intervention 'works' in the sense that it alters the course of a disorder (Olswang, 1990; Law, 1997).

[5] Ward's study (1994) consisted of three groups, each with their own matched controls. On the basis of the information available to us, it was not possible to neatly combine the three groups, and combining the effect sizes would have resulted in a marked reduction in the functional sample size and a consequent reduction in statistical power. For this reason, Group 2 (n = 13), who had receptive and expressive language difficulties but no associated listening difficulties were considered a replication and included as a separate study. The numbers in the third group of children with specific expressive difficulties were so low (n = 4) as to preclude derivation of meaningful effect sizes and as indicated above, could not be combined with one of the other groups.

[6] There was very little variability in the reliability scores. The mean was 7.10 (SD 0.53) from a maximum possible score of 9, so the scores are not reported here.

[7] Five of the studies on average had less than ten subjects per group.

[8] Ward (1994) did not report the results of any statistical analysis but provided sufficient data for a comparison of post-test means to be carried out.

[9] The interpretation of outcomes here is further complicated by differential attrition rates across groups (Whitehurst: personal communication, 1997).

[10] Stevenson *et al* (1982), Ward (1994, Groups 1 and 2), and Whitehurst *et al* (1991) report the use of standard scores in their analyses.

[11] The marked variation in sample size between studies posed problems for the assumption of homogeneity of variance required by conventional ANOVA techniques. Hedges and Olkin's (1985) procedures for weighted ANOVA were used throughout. The procedures have an advantage in that they provide a means not only of comparing variance between groups (Q_b which is analogous to an omnibus F-test for between-groups differences but without the assumptions) but also a test of the hypothesis that the data come from the same population using a statistic, Q_w, which provides a test for within-group variation in effects and which has an approximate χ^2 distribution with k-1 degrees of freedom (df), where k is the number of effect sizes.

[12] See the example in appendix 6 where study design and indirect versus direct treatment are confounded.

[13] Both of these reviews included a wide range of study quality including non-controlled studies.

[14] In this respect the analysis here differs from that of Scruggs and co-workers (1988) who included multiple measures.

[15] A weighted least-squares meta-regression analysis (Hedges and Olkin, 1985) was carried out to examine the relationship between effect sizes from the group designs (109 in total) and child characteristic (e.g. age), and treatment variables (e.g. type of intervention, duration, frequency). However, the resulting model was not well-specified (Hedges and Olkin, 1985) and the results are not reported here.

Chapter 6

The accuracy of screening procedures

Review questions

- What is the reported range of productivity figures for screening procedures?
- Do the test characteristics differ across the population ages screened?
- What are the characteristics of the more accurate methods of screening for speech and language delays?
- What evidence is there that screening measures identify more cases than are already identified through self-referral?
- Can the introduction of screening for speech and language delays be shown to be feasible at different ages in terms of cost, manpower, uptakes, and identification rates?

Inclusion/exclusion of literature

This review focuses only on the screening procedures that are adequately described and can therefore be replicated, and which have been evaluated for their accuracy. Full details of the inclusion/exclusion criteria are given in appendix 4.

Relevance Studies of the application of replicable screening procedures to normal and clinical populations within the 0–7 year age range. The screening procedures target speech and language skills.

Participants Information about the number of participants in each group at both the screening and the diagnostic stage.

Outcomes 2×2 tables could be derived, allowing computation of productivity figures, or statement of productivity figures.

Designs (a) Randomised screen/no screen studies
(b) Two-stage screen/diagnostic studies of representative populations
(c) Two-stage screen/diagnostic studies of clinical populations.

All the screening procedures examined explicitly refer to speech and language skills. Similarly, the reference tests adopted as benchmarks for the screening procedures all explicitly access speech and language skills. Thus, for screening procedures that cover a wide range of development, the element addressing speech or language skills had to be separately validated against a stated speech or language reference test. (Outcome restricted to developmental status *per se* was not acceptable evidence for this review.) The aim was to identify evaluations of screens rather than summarise data from test manuals. No attempt has been made to evaluate the face validity of screening procedures or to make recommendations about the value of specific tests. For this the reader is referred to Glascoe and co-workers (1990) and Sturner and co-workers (1994). The intention in this review is to describe characteristics of screens, without necessarily making clinical recommendations for the use of a particular screen.

Diagnostic measures and their evaluations are not included in the review. Decisions regarding the distinction between a screen and a diagnostic test were based on:

- the description of the potential users of the test; tests used only by speech/language clinicians were considered to be diagnostic in character
- the description of the purposes of the test; tests informing therapy content decisions/differential diagnosis were considered diagnostic
- the time needed for the test; although this can vary widely for screening procedures, in the absence of other information, a test taking longer than 30 minutes was not counted as a screening instrument.

Studies reporting correlation data, and not classification data, were excluded. That is, where the screen scores were correlated to the reference test scores, but no data given that could generate productivity figures, the study was excluded. (Further discussion of the correlation and the classification approaches can be found in Lichtenstein and Ireton, 1984.) The outcome of each validation study was reduced to the 2×2 classification table from which productivity figures could be derived. Where these were not available the review team worked from the stated productivity figures or percentages.

Within the included evaluation studies, the sample could reflect a general population, or include

TABLE 16 *A 2 × 2 table used to express classifications*

	Reference test positive 'abnormal'	Reference test negative 'normal'	Total
Screen-positive 'abnormal'	a True-positives	b False-positives	a + b
Screen-negative 'normal'	c False-negatives	d True-negatives	c + d
Total	a + c	b + d	a + b + c + d = n

known speech and language delay cases. This follows the clinical development of screening procedures, which may well be tested on known cases before being evaluated further on a general sample. The external validity of the clinical studies is indeed compromised by their sampling, and this was reflected within the quality rankings assigned (see below).

Framework for the analysis

Screening as a process is intended to separate out the children who need further investigation of their speech and language skills from those with normally developing speech and language. In doing this, a screening procedure ('**screen**') classifies children as 'possibly abnormal' or 'possibly normal'. An acceptable screen is one which leads to a classification decision which is corroborated by a more 'in-depth' assessment, with low numbers of children incorrectly considered possibly abnormal or possibly normal. Resources can then be concentrated on the children failing the screen. Thus, the essence of an evaluation of a screen is its comparison with another assessment, showing acceptable limits of incorrect classification. The evaluation process is also referred to as **validation**.

The in-depth assessment used in the evaluation is referred to as the **gold standard**, or r**eference test**. It is expected to be a standardised clinical **diagnostic test**, of known validity. In the evaluation process, all the children given the screen, or a subsample, are also given the reference test. Each child is classified as normal or abnormal on both the screen and the reference test. For example, all scores below 1 SD from the mean on the reference test could be called 'abnormal'; while all children failing more than four items on a screening checklist may be deemed 'abnormal' or screen-positive. The decision level on each test is termed the **cut-off**. The results give rise to a comparison of classifications, which are expressed in a 2 × 2 table (*Table 16*). This table may be read horizontally or vertically. Read horizontally it provides population-specific figures allowing interpretation of the adequacy of the measure in a given population. Read vertically it provides population-independent data, which can be used to make some degree of comparison across studies. Statistics derived from the 2 × 2 matrix are collectively known as **productivity figures**. In the present review, three key population-independent productivity figures are reported from the studies reviewed, namely sensitivity, specificity, and the LR (see *Box 3*). One population-specific figure is reported, namely positive predictive value (PPV).

These indices of a screening test have immediate clinical application. Moreover, they are not subject to variance if the target population has a higher prevalence of speech and language delay. The LR, being derived from the sensitivity and specificity,

BOX 3 The key productivity figures

Sensitivity refers to the proportion of clinical cases (with delayed speech and language development) correctly classified by the screening procedure. It is given by a/(a + c).

Specificity refers to the proportion of normal cases (children with normal speech and language development) correctly classified by the screening procedure. It is given by d/(b + d).

LR refers to the likelihood of a positive screen result. It is given by [sensitivity/(1 – specificity)]. It expresses the odds that a given cut-off level of a screening test would correctly identify a child who has 'true' speech and language delay.

PPV of a screen focuses on the diagnostic outcomes for a positive screen result, being the proportion of screen positives who are true cases. It is given by a/(a + b). Where it is of interest to know the PPV of a screening test on a second population, it can be derived from the screening test result (in terms of LR) on population A, together with the estimated prevalence of the disorder in population B.

is sometimes taken as a 'shorthand' for the test's performance. A high specificity usually leads to a high LR; for two screens of equal sensitivity, the one with the higher specificity will have the higher LR. In general, a high LR is desirable.

A given screening test is unlikely to be completely accurate and it is hard to maximise sensitivity and specificity together. The accuracy of a given screening measure is always relative to the reference test chosen and to the expectations regarding adequate levels of sensitivity and specificity. For example, the figure of 0.8 could be considered acceptable accuracy for both dimensions. This would demand that, of the children tested, a maximum of 20% of the clinical cases are erroneously deemed 'normal', and that a maximum of 20% of normals are wrongly classified as 'abnormal/clinical'. With notional prevalence figures of 10 cases in every 100, the numbers would be 8 of those 10 clinical cases correctly classified, with 72 of 90 normals correctly classified. In this example, the LR would be 0.8 divided by 0.2, or 4.0. A screen which sought to maximise sensitivity may have productivity figures of 0.9 (sensitivity), 0.7 (specificity) and 3.0 (LR). By contrast, a screen maximising specificity could have figures of 0.7 (sensitivity), 0.9 (specificity) and 7.0 (LR).

The significance of different outcomes following a screen will determine the levels of accuracy demanded. For a disorder with a prevalence of 10%, a sensitivity of 0.7 would suggest that after screening 1000 children, 70 of the 100 cases are identified. The clinical significance and costs of missing the other 30 true cases (and indeed, the significance of misclassifying some normals) should underlie whether or not 0.7 sensitivity is considered acceptable. Thus, the severity of the screened condition and its prevalence impact upon the levels of sensitivity and specificity tolerated. One element in the consideration of the costs of a screening programme is how many true cases are identified.

When developing a screen, researchers and clinicians may consider trying different screen cut-off scores, against the same reference test, in order to see how the corresponding figures of sensitivity and specificity vary. This generates a set of data pairs of sensitivity and specificity for each screen cut-off. These can be displayed as a receiver operating characteristic curve, (**ROC curve**), in order to visualise the optimisation of sensitivity and specificity.

A screen tested on one population gives population-specific information about that screen. Any one of the productivity figures quoted will give some indication of the accuracy of a screening measure evaluated in this way. However, only PPV will vary according to the prevalence of the target disorder in the population concerned and so cannot readily be used to compare screening tests. A screen validated across a variety of populations, against a consistent reference test, provides much stronger data for the validity of the screen and for its general application. Comparative information can also be obtained if more than one screen is tested simultaneously on one population, for then the relative performance of the screens can be described. By extension, it is possible to test several screens simultaneously across several populations.

The validation process may be conducted concurrently or predictively. In a **concurrent validation**, the reference test is given at the same time as the screen, or within a short interval. In a **predictive validation**, the reference test is used at a later point in time, usually more than 6 months later, to establish the predictive power of the earlier screen. The descriptive statistics used are the same. Here the studies reviewed are concurrent validations. For a discussion of predictive issues in the context of the natural history of speech and language delay, see chapter 4.

The screens reviewed here are either **monophasic**, in which they focus exclusively on speech or language abilities, or **multiphasic**, in which they tap a range of developmental skills, including speech or language. The method of assessment is usually either by direct testing of the child, or by asking a parent or carer certain questions, or by observation of the child.

Review of the data

The evidence reviewed here consists of screens applied to whole populations, with the performance of the screen being evaluated in terms of its productivity figures. (A complete summary of the papers can be found in appendix 5.)

In the 45 papers reviewed, some authors used more than one cut-off on the reference test or on the screen. Some authors also quoted productivity figures for subsets of the sample by age groupings. This generated 85 data sets of productivity figures, which will be referred to as studies. A total of 19 data sets were from research designs applying two screening tests to one population. Sixty-six data sets were reported for single-screening tests applied to one population.

There were no studies of multiple-screening procedures being evaluated simultaneously over a variety of populations. Thus, most of the studies evaluated one screening test on one population. Quantitative combination of these results is not feasible. Although different authors have reapplied a given screening test to a fresh population, they used a different reference test, making comparisons unwarranted. Instead, descriptive comments can be made about the ranges of sensitivity and specificity for children of different ages, and for the screening method used. First, an analysis is given for the studies with more than one screen, or more than one population.

One screen/multiple populations

The strongest analysis of screening procedures is that of combining several studies which have looked at the performance of one procedure over several populations (with a consistent reference test). The methodology for this has been developed into a meta-analytical technique called the **summary ROC curve** (Irwig *et al,* 1994 and 1995). However, the data identified in the present review were not of a type to permit the application of such a technique.

The data on the Fluharty Preschool Language Screening Test are from three studies (Illerbrun *et al,* 1985; Sturner *et al,* 1993a Studies 1 and 2). The reference tests used in these studies were similar, but not identical (*Table 17*).

The samples in these studies were drawn from two US States, with Sturner's populations coming from a rural county. All studies targeted children aged from 5 years. Although specificity is comparable across the studies, the sensitivity estimates are very varied. So while the Fluharty test in different situations is specific in excluding true normals, it may miss 35–83% of the true cases (*Table 17*).

The data on the Sentence Repetition Screening Test (*Table 18*) comes from Sturner and co-workers (1993b and 1996). Sturner applied the same screen and reference testing procedures to two populations that differed in age range. However, the samples are drawn from the same geographical area, a rural county in the USA. For both the speech and the language figures, sensitivity improves with the older population (Population 1). Specificity is high across the ages tested and the area of language tested. Taking LR as a test indicator, speech screening has the higher LR at the younger age, while language screening has a higher LR at the older age range.

Both the Fluharty and the Sentence Repetition Screening Test have been developed in the USA and are not used by UK speech and language therapists.

Multiple screens/one population

Eight papers were reviewed which dealt with more than one screening procedure on one population. Of these, five papers evaluated two procedures that both met the screen criteria used for this review (*Table 19*).

These data allow comparison of screens within one population; the higher LRs suggest the more discriminating screens. Also, within one population, PPVs can be compared to show the relative performance of a screen in locating true cases. In *Table 19*, the multiphasic tests (Denver, Battelle, Developmental Profile II) were tested on clinical or mixed populations, but the monophasic tests were used with normal populations only. The multiphasic tests have LRs in the range of 1.12–3.3, but the range for the monophasic tests is wider, at 1.04–28.17. It is interesting to note that even when sensitivity and specificity are similar for two screens, the LR and PPV can still reflect differences between the screens.

TABLE 17 *Studies of the Fluharty Preschool Language Screening Test*

Study	Reference test	Age (months) [n*]	Sensitivity	Specificity	LR
Illerbrun *et al,* 1985	Combined TOLD/TACL/CELI	68–77 [136/136]	0.65	0.94	10.77†
Sturner *et al,* 1993a Study 1	TOLD (language figures)	53–68 [279/378]	0.38	0.85	2.53
Sturner *et al,* 1993a Study 2	TACL (language figures)	55–69 [421/533]	0.17	0.97	5.67

* *The population size (n) is given as the sub-sample size/total sample size*

† *LR calculated on full values of sensitivity and specificity before their truncation to two decimal places*

TOLD = Test of Language Development; TACL = Test of Auditory Comprehension of Language; CELI = Carrow Elicited Language Inventory

TABLE 18 *Sentence Repetition Screening Test*

	Reference test	Age (months) [n*]	Sensitivity	Specificity	LR†
Speech					
Population 1	Arizona Articulation Proficiency Scale	63–96 [78/382]	0.74	0.92	9.69
Population 2	Arizona Articulation Proficiency Scale	54–66 [76/343]	0.57	0.95	11.40
Language					
Population 1	Illinois Test of Psycholinguistic Abilities Bankson Language Screening Test	63–96 [78/382]	0.76	0.92	9.41
Population 2	Illinois Test of Psycholinguistic Abilities Bankson Language Screening Test	54–66 [76/343]	0.62	0.91	6.90

* *The population size (n) is given as the sub-sample size/total sample size*
† *LR calculated on full values of sensitivity and specificity before their truncation to two decimal places*

Study data combined

The range of sensitivity, specificity and LRs for all the studies summarised in the preceding two sections together with the studies dealing with one screen on one population are shown in *Table 20*, with a distinction drawn between those evaluations on samples taken from a normal population, and those taken from populations including known speech and language delay cases.

The extent to which the productivity figures are a function of study quality is important. Accordingly, the relationship between the quality ranking and the productivity figures was examined. The method of ranking the quality of studies is given in appendix 8. The rankings were on a scale of 0–33, a grade of 'high' was assigned to scores equal to or higher than 22. Across all the studies there was a negative correlation between study quality and LR ($r = -0.23$, $p < 0.05$) suggesting that lower quality studies tend to generate higher LR. This relationship was particularly strong ($r = -0.34$, $p < 0.05$) for the screening studies which used normal samples even though, as a group, the normal sample had a higher mean LR. The quality ranking was also significantly negatively correlated ($r = -0.35$, $p < 0.005$) with the sensitivity but not with specificity across all groups, a finding that appears particularly robust in the studies which used normal samples ($r = -0.48$, $p < 0.001$) but not those that used clinical samples. In practical terms, this suggests that it is generally easier to state who is 'normal' than it is to identify true cases. Where mixed clinical/normal populations are used the number of cases is effectively over-sampled, thus increasing the chance of identifying true cases. Finally, it is apparent that the high-graded studies tend to have a significantly higher specificity than sensitivity ($t = 4.41$, $p < 0.001$), a relationship that disappears for the lower quality studies. This suggests that case status may be a particular problem where normal samples are used.

It is worth noting from the ranges reported above that there are studies which report extremely low figures for sensitivity and specificity. That such figures are reported at all can probably best be explained by looking at other aspects of the papers, such as the estimated PPV. The researchers and clinicians may pay more attention to the accurate identification of cases in their own populations (i.e. maximising PPV) than whether their measure would misclassify cases in the overall population.

Range of sensitivity and specificity by age

To determine whether sensitivity, specificity and LRs vary according to age, each study was assigned to one age band:

- **under 2 years** included studies with children all under the age of 2 years
- **2–3 years** included studies with children up to the age of 3 years (including some infants)
- **3–5 years**
- **≥ 5 years**
- **0–7 years** included studies with children across the pre-school/school range.

The ranges of productivity figures are given in *Table 21* for studies with normal samples, and in *Table 22* for studies with clinical samples. As in *Table 20,* means are given for descriptive purposes only.

TABLE 19 *Studies comparing screens on one population*

Study	Screening procedure	Criteria for language delay (reference test, cut-off)	Age (months) [n*]	Productivity figures			
				Sensitivity	Specificity	LR	PPV
Allen and Bliss, 1987	Fluharty Preschool Language Screening Test	Sequenced Inventory of Communication Development; receptive language at least 12 months below CA, or expressive language at least 12 months below CA	36–74 [182/182]	0.60	0.80	3.15	0.33
	Northwestern Syntax Screening Test			0.92	0.48	1.79	0.22
German *et al,* 1982 (clinical population)	Revised Denver Developmental Screening Test	Sequenced Inventory of Communication Development; scored as in manual	Mean 41.7 [84/84]				
	Expressive						
	Conservative			0.92	0.49	1.79	0.71
	Liberal			0.96	0.14	1.12	0.61
	Receptive						
	Conservative			0.95	0.45	1.74	0.63
	Liberal			0.98	0.14	1.14	0.53
	Developmental Profile II						
	Expressive						
	Conservative			0.92	0.72	3.30	0.81
	Liberal			0.98	0.42	1.68	0.69
	Receptive						
	Conservative			0.93	0.62	2.44	0.71
	Liberal			1.00	0.36	1.56	0.61
Glascoe and Byrne, 1993 (mixed population)	Battelle Development Inventory Screening Test	Fluharty Preschool Speech and Language Screening Test; 3 sub-tests failed **or** Vineland Adaptive Behavior Scale communication quotient more than 1.5 SD below IQ	7–70 [89/89]	0.78	0.70	2.63	0.40
	Denver Developmental Screening Test II			0.73	0.76	3.02	0.43
Stokes, 1997	Parent questionnaire a) with comprehension item	Reynell Developmental Language Scales; –2 SD **or** language sample analysis; at stage I/II/III of syntax and/or phonology development	34–40 [398/398]	0.78	0.91	8.33	0.56
	b) without comprehension item			0.78	0.95	17.21	0.72
	Nurses developmental screen			0.77	0.97	28.17	0.78
Sturner *et al,* 1996	Sentence Repetition Screening Test	**Language** ITPA auditory reception and auditory association sub-scales; Bankson Language Screening Test; both at less than 30th centile	54–66 [76/343]	0.62	0.91	6.90	0.44
		Speech Arizona Articulation Proficiency Scale; less than 15th centile		0.57	0.95	11.40	0.75
	Speech and Language Screening Questionnaire	Language		0.59	0.43	1.04	0.12
		Speech		0.68	0.89	6.18	0.66

* *The population size [n] is given as the sub-sample size/total sample size. ITPA = Illinois Test of Psycholinguistic Abilities*

TABLE 20 *Summary of mean productivity figures for studies with normal samples and those with clinical samples*

	Minimum	Maximum	Mean – all studies	Mean – high-grade studies only
Normal samples				
No. of studies			51	23
Sensitivity	0.17	1.00	0.78	0.65
Specificity	0.43	1.00	0.88	0.88
No. of studies*			45 (44)	22
LR†	1.04	1158.50 (59.10)	37.30 (11.80)	12.10
Clinical samples				
No. of studies			34	5
Sensitivity	0.30	1.00	0.82	0.72
Specificity	0.14	1.00	0.72	0.88
No. of studies*			31	4
LR	1.12	36.90	6.00	8.50

Note: The means are used descriptively to anchor the range of a given statistic. They are not used to suggest that an 'average' screen exists which would generate the mean productivity figures. LR: whenever specificity = 1, LR is a ratio with zero denominator. In such cases, LR is not well defined and cannot be quoted. Thus, the number of studies with well defined LR is lower than the total number of studies.

* *Total number of studies with LR well defined: 76*

† *Figures in parentheses represent the set with the one outlying LR of 1158.5 excluded*

TABLE 21 *Range of productivity figures for screen evaluation studies by age band of the validation sample: normal sample studies*

	< 2 years	2–3 years	3–5 years†	≥ 5 years	0–7 years
No. of studies	4	15	12 (11)	3	17
No. of screens*	3	5	7 (6)	2	6
Sensitivity range	0.50–0.92	0.67–1.00	0.77–1.00	0.69–0.90	0.17–0.97
Sensitivity mean	0.74	0.91	0.88 (0.87)	0.80	0.60
Specificity range	0.67–0.92	0.75–1.00	0.78–1.00	0.61–0.79	0.43–0.98
Specificity mean	0.86	0.93	0.92 (0.91)	0.71	0.85
No. of studies	4	10	11 (10)	3	17
LR range	2.07–11.80	4.00–45.00	3.60–1158.50 (3.60–59.10)	2.06–4.35	1.04–54.90
LR mean	7.40	15.90	120.20 (16.40)	3.00	9.40

Note: The means are used descriptively to anchor the range of a given statistic. They are not used to suggest that an 'average' screen exists which would generate the mean productivity figures. LR: whenever specificity = 1, LR is a ratio with zero denominator. In such cases, LR is not well defined and cannot be quoted. Thus, the number of studies with well defined LR is lower than the total number of studies.

* *Some screens appear at more than one age band, if validated separately. Denver versions (DDST; DDST-R; DDST II) taken as one screen*

† *Figures in parentheses represent the data set with one low-graded study with outlying LR removed*

The small number of studies in the under 2 years and for 5 years and over categories make it difficult to examine the relationship between mean LR and age. However, the mean LR does appear to show some improvement between the under 2 and the 2–5 years groups, but then appears to tail off after 5 years (mean LR rising from 7.4 to 16.4). A similar pattern is seen in the data from clinical samples. Thus, the data suggest identification may be more accurate for screens that exclusively target children in the 2–5-year age range.

Range of sensitivity and specificity by screen method

Screen studies were compared according to their method of assessment (i.e. direct assessment, parent or carer report, or observational approaches). Screens using a mixed method were classified by their principal approach. Only two studies used an

TABLE 22 *Range of productivity figures for screen evaluation studies by age band of the validation sample: clinical sample studies*

	< 2 years	2–3 years	3–5 years†	≥ 5 years	0–7 years
No. of studies	2	9	5	None	18
No. of screens	2	5	3	None	6
Sensitivity range	0.83–1.00	0.50–1.00	0.58–1.00		0.30–1.00
Sensitivity mean	0.92	0.85	0.81		0.79
Specificity range	0.55–1.00	0.69–0.97	0.62–0.98		0.14–1.00
Specificity mean	0.78	0.86	0.77		0.62
No. of studies	1	9	5	None	16
LR range	2.30	3.17–20.80	1.65–36.90		1.12–14.90
LR mean	2.30	9.50	10.30		3.00

Note: The means are used descriptively to anchor the range of a given statistic. They are not used to suggest that an 'average' screen exists which would generate the mean productivity figures. LR: whenever specificity = 1, LR is a ratio with zero denominator. In such cases, LR is not well defined and cannot be quoted. Thus, the number of studies with well defined LR is lower than the total number of studies.

observational method, and therefore, the analysis here focuses on the 83 studies that used parent report or direct assessment. The productivity figures are given in *Tables 23* and *24*.

Mean sensitivity and specificity figures from the normal samples suggest that the two methods do not show distinct differences; parent/carer report in the normal samples appears to be more sensitive than direct testing methods, at the cost of being a little less specific, but these differences are not apparent in the clinical samples, and none of them reach statistical significance. It is noted that some direct assessment methods have very low sensitivity. The higher specificity of direct methods is reflected in the mean LR for these methods. A similar pattern is seen in the data from clinical samples; the ranges of sensitivity and specificity are more varied for direct assessment methods, but the overall LR is comparable for the two approaches.

Characteristics of screens with higher productivity figures

The data were examined to identify the screens in each age band with higher levels of sensitivity and specificity. These are shown in *Table 25*. Most of these data come from studies with normal populations. The few high-performing screens from clinical populations are indicated.

Thus, the younger age groups (up to 5 years) are well represented, with screens performing at high levels of sensitivity and specificity. In very young children (up to 2 years) parent reporting is the method used. After the age of 2 years screens using either parent report or direct assessment of the child can achieve the stated levels of sensitivity and specificity. There are few studies in this review of

TABLE 23 *Range of productivity figures for screen evaluation studies using direct child assessment compared with parent or carer report: normal sample studies*

	Parent/carer report*	Direct assessment
No. of studies	23	27
No. of screens	10	9
Sensitivity range	0.50–1.00	0.17–1.00
[mean]	[0.81]	[0.76]
Specificity range	0.43–1.00	0.48–1.00
[mean]	[0.86]	[0.90]
No. of studies	21 (20)	23
LR range	1.04–1158.50 (1.04–42.00)	1.79–59.06
LR mean	63.80 (9.10)	13.50

* *Data in parentheses represent data set with one low-graded study with outlying LR removed*

TABLE 24 *Range of productivity figures for screen evaluation studies using direct child assessment compared with parent or carer report: clinical sample studies*

	Parent/carer report	Direct assessment
No. of studies	12	21
No. of screens	5	7
Sensitivity range	0.50–1.00	0.30–1.00
[mean]	[0.85]	[0.80]
Specificity range	0.36–1.00	0.14–1.00
[mean]	[0.73]	[0.70]
No. of studies	11	19
LR range	1.56–17.60	1.12–36.90
LR mean	5.90	6.00

TABLE 25 *Characteristics of screens with higher levels of sensitivity and specificity*

Productivity figures	< 2 years	2–3 years	3–5 years	5–7 years	0–7 years
Sensitivity > 0.9; specificity > 0.9 Means: sensitivity 0.98 specificity 0.94 LR 24.8	WILSTAAR	SKOLD LDS Levett-Muir ELM* Hackney*	SKOLD Uppsala language screen Stevenson screen TPSI*	None	DDST language sector
Sensitivity > 0.9; specificity > 0.8 Means: sensitivity 0.97 specificity 0.93 LR 21.8		LDS SKOLD			ELM* DDST language sector*
Sensitivity > 0.8; specificity > 0.9 Means: sensitivity 0.94 specificity 0.94 LR 23.3	CLAMS (receptive validation)	CLAMS (expressive validation) LDS	Rigby speech screen Stevenson screen		
Method of assessment	Parent report	Direct (3) Parent report (3)	Direct (4) Parent report (1)	N/A	Direct (1) Parent report (1)

* *Clinical sample*

CLAMS = Clinical Linguistic and Auditory Milestone Scale (Clark et al, *1995)*
DDST = Denver Developmental Screening Test (Dodge, 1980)
ELM = Early Language Milestone Scale (Coplan et al, *1982)*
Hackney = Hackney Early Language Screening Test (Dixon et al, *1988)*
LDS = Language Development Survey (Rescorla, 1993)
Levett-Muir = Levett-Muir Screening Test (Levett and Muir, 1983)
Rigby Speech Screen = Speech Screening Test (Rigby and Chesham, 1981)
SKOLD = Screening Kit of Language Development (Bliss and Allen, 1984)
Stevenson Screen = Stevenson Screening Test (Stevenson and Richman, 1976)
TPSI = Texas Preschool Screening Inventory (Haber and Norris, 1983)
Uppsala = Uppsala general language screen (Westerlund and Frylmark, 1997)
WILSTAAR = Ward Infant Language Screening Test; Assessment, Acceleration and Remediation (Ward and Birkett, 1994)

TABLE 26 *Variety of cut-off score definition methods for reference tests*

LA	SDs below the mean	Standard scores	Percentile scores	Other methods
LA < CA – 6 months LA < CA – 12 months	–1	< 80	< 10th, 15th, 25th, 30th	Language sample analysis at Stage III
LA < 0.66 CA LA < 0.70 CA	–1.5 –2	< 85		Three sub-tests failed 'Severe' score

the older range (5–7 years); the study by Culatta and co-workers (1983) was the highest performer, with figures of 0.9 for sensitivity, 0.79 for specificity.

Reference tests

The function of the reference test is to determine true case status. However, the studies reviewed showed considerable variety in how this case status was defined. Not only is a range of reference tests used but there is also variation in the cut-off score employed. The cut-off scores were defined in terms of SDs, standard scores, percentiles and discrepancy of CA to a derived LA. This is summarised in *Table 26*. Further, studies using a given test (e.g. the Reynell Developmental Language Scale) specified cut-offs differently (e.g. –1 or –2 SDs,

or language delay of 6 months relative to CA). Thus, comparison of screens against a set reference test has not been possible.

The impact of how case status is defined is seen most clearly in the 'implied prevalence' of an evaluation study. This is defined as the proportion in the total sample falling below the cut-off score on the reference test. (In terms of the 2×2 table it is given by a + c /n). Of the 17 studies with high-quality gradings, which reported data from normal populations (or those which stratified their samples in such a way as to reflect their original populations), and for which 2×2 data could be reconstructed, prevalence ranged from 5% to 23.5% with a mean of 14.0%. This suggests that many studies build in prevalence rates which exceed most existing prevalence estimates, presumably with a view to exercising caution with more liberal cut-offs. If this is the case, the use of a liberal diagnostic may underly the relatively lower sensitivity rates seen for normal sample, high-quality studies. This issue of case definition is crucial to the evaluation of any screening procedure, and to any intervention outcomes, and deserves more attention.

Screening programmes

Only one study has specifically addressed the value of screening procedures relative to self referral (Drillien *et al,* 1983). This did not meet the inclusion criteria for this review. However, of the 125 children referred to speech and language therapy services all but three came through the screening system. This issue necessarily depends on both the sensitivity of the identification procedure and on the perceived need for support services of this nature by parents and other professionals. It is likely that this perceived need will vary according to a range of local variables such as the level of local services, the motivation of professional staff, and the degree of parental knowledge of child development in general and concern about early speech and language development in particular.

It is also important to consider the role played by primary prevention. Health visitors in the UK provide advice on child-care practice and encourage parents to speak to young children in an appropriate manner. Parents then choose whether they need to modify their own behaviour, thus effectively identifying their child themselves as a potential case, in collaboration with their health visitor. Again no studies have explicitly addressed the issue of whether primary prevention in the sense of advice of this sort could effectively replace formal screening procedures. However, there is evidence presented here that parental judgement as to whether their child is delayed may be at least as accurate as a test procedure specifically designed to elicit behaviour from the child and carried out by a third party.

No studies examined the relative value of formal screening of the type described here and identification based on a decision made by primary-care professionals in collaboration with parents. Such a project is currently under way in East London but will not be reported until 1999 (Laing, Law, Logan and Lewin).

Costs

The aim of screening and intervention is ultimately to reduce the prevalence of a given condition and thereby reduce costs to the services concerned over the long-term. It is not possible at this stage to comment on the costs and benefits of screening for speech and language delays because no studies have explicitly compared screened and unscreened populations in terms of their subsequent use of services, though a study is currently underway in Holland by a team from the Department of Public Health at Erasmus University, Rotterdam,and data are expected in 1999 (Konig, Reep, van Agt, Polder, Korfage and van der Maas). Such investigations need to go beyond mechanical costing of materials and staff time (Bricker *et al,* 1988) to look at longer-term benefits such as achievement in school, perceived benefit on the part of the child or their parents, school, etc. To date the lack of clarity in outcomes has been a problem as has the lack of appropriate economic indicators in this area, equivalent to the quality-adjusted life year.

A screening procedure's yield is one element of cost analysis. This refers to the number of true cases identified relative to the total number of cases screened (or a / a + b + c + d) and is closely associated with the prevalence of the condition concerned. A high prevalence is likely to lead to a high yield. The range of yields was estimated for the 17 normal sample studies allocated a high-quality ranking, for which 2×2 data could be reconstructed. The range was 3.6–12.6%, with a mean yield of 9.3%. Butler (1989) indicates that yields in excess of 5% are very good in screening terms and this would suggest that on productivity figures alone many of the screening procedures reported here offer acceptable rates of detection and could therefore be considered cost-effective.

The screening process must also be concerned with the utilities attached to different outcomes.

Concern has been expressed elsewhere that the psychological and, indeed, the health costs of screening for some conditions may exceed the benefits to be gained from such a process (Stewart-Brown, 1997). In particular, there may be costs associated with relatively high false-positive and false-negative rates. The data presented in chapter 4 provide convincing evidence that intervention is likely to lead to a more favourable outcome than no treatment for a child with language delay, at least in the short term. This evidence effectively raises the costs associated with not treating true cases as a result of the low sensitivity associated with a given screen. The converse of this, inappropriately intervening in cases where the child's difficulties are likely to resolve spontaneously, is a more complex issue. Those providing such intervention are likely to argue that the effects would be benign. However, there may also be costs associated with lowered expectations of the child and increased anxiety that have yet to be investigated in the context of screening for speech and language delays. There will also be costs associated with inappropriate use of staff time.

Manpower

No studies have specifically examined the staffing issues associated with screening for speech and language delays. If health visitors and GPs assess children and monitor development, a screening programme may be practical. If this type of service is withdrawn, however, there is no other group of professionals with sufficient numbers or coverage to replace it. Likewise for screening to be a practical possibility there must be sufficient services available to support the children and their families. At one level this refers to routine nursery and school support. At another, there must be sufficient speech and therapy provision and psychological support to assist children whose speech and language needs are greater than those of their normal peers. The introduction of a systematic screening procedure for speech and language delay is also likely to increase referrals to agencies dealing with associated behaviours, for example, community paediatricians, child psychiatrists and audiologists.

Uptake

There has been extensive discussion of uptake or coverage of screening (Butler, 1989) but little of it relates specifically to speech and language screening. In most cases the screen evaluations reported in this review relate to the discriminatory power of the procedure itself, but pay little attention to the numbers that do not present for screening. As indicated, there is a lack of data on population-based applications of a screening procedure. The Drillien and Drummond study (1983) covering the Dundee area reported high uptakes of general developmental screening, ranging from 95% at 8 weeks of age to 82% at 3 years. This was in the context of a rigorous system for achieving contacts. By contrast, the attendance rate for developmental screening in which the speech and language element was a component, was much lower in a deprived inner city area (54.2%) in a population of children aged 2.6 years (Law, 1994). In the latter study no attempt was made to trace those children who did not attend for developmental surveillance. This level of non-attendance would clearly pose problems for population screening.

For population screening to be truly effective the majority of cases must be identified. High non-attendance rates are likely to invalidate the process. This poses something of a paradox for the providers of services to children with speech and language delays. Given that it is difficult to identify such delays at a truly pre-symptomatic stage it is likely that screening will only become appropriate once particular milestones have been passed by the majority of cases. Attendance at general developmental screening tends to drop off after 2 years of age and if this drop is too large it will threaten the validity of the procedures concerned.

Identification rates

Sometimes the accuracy of the screening procedure is expressed in terms of the hit rate or classification rate derived from the sensitivity and specificity, namely $(a + d / n)$. This represents the total number of children correctly classified by the screening procedure. As a single figure this does not indicate whether the test is more specific or more sensitive. Of the normal sample studies allocated a high quality grading, 2×2 data could be reconstructed for 17 studies. This gave a range for their identification rate of 72.5–98%, with a mean of 86%, indicating highly acceptable rates.

Summary

- The present data set included screening measures that could be used in child health surveillance and for which the data were reported in terms of the tests' capacities to classify relative to a gold-standard measure. Studies used normal populations and mixed clinical/normal populations.
- A number of screening tests have been shown to have adequate specificity and sensitivity. In high-prevalence conditions it is more acceptable to have higher specificity and lower sensitivity. The reverse would be true of low-prevalence

high-severity conditions. Given the relatively high prevalence discussed in chapter 3, the range of productivity figures reported for the better quality studies is expected. It is noted, however, that those authors researching screens do not explicitly link their reference-test criteria to estimates of prevalence. In many cases the implied prevalence adopted in the screening studies exceeded that which might be estimated for the population as a whole.

- Studies with higher quality ranking showed higher specificity than sensitivity, suggesting that it is easier to be accurate in identifying children who are not cases than it is to identify those who are cases.
- The LRs (adopted here as a descriptive measure of screening productivity) reported are inversely related to study quality across all samples.
- Most of the literature reviewed does not report data that could allow manipulation of cut-offs (relative to different definitions of case status) using techniques such as the ROC analysis. This makes it impossible to judge the extent to which researchers have sought to optimise cut-offs on the procedures concerned.
- Parental report and child-focused testing result in comparable productivity figures. This suggests that parents may be as good as screening tests at identifying children with speech and language delays, though this does depend on eliciting parental judgement appropriately.
- The reported accuracy of screens varies according to the age of the child. The reviewed data appear to show higher mean LRs for screens in the 2–5 years age range. However, it is not possible to specify at which age the introduction of a screening procedure would be most appropriate because no study has yet examined the relative value of identification at different ages. Similarly, it is not possible using the natural history data, to specify whether one screen would be sufficient to identify all children who are likely to have persistent problems. For example, given the high level of spontaneous remission of expressive delays reported at 2 years, the introduction of a single measure with coarsely grained cut-off points (e.g. a parent report of vocabulary) at this stage in a child's development is unlikely to be sufficient.
- The issue of screening coverage has received little attention in the present data set, though evidence does suggest that this remains an obstacle to the effective implementation of universal screening programmes. This is especially relevant for screening for early speech and language delays because the period for which the reported screens are most accurate (2–5 years) is generally perceived to coincide with declining attendance for child health surveillance.
- It is not possible to specify which of the screening procedures described would be a best choice because there are very few data that would allow comparison of performance across screening measures.
- The considerable variation in the definition of case status using a range of gold standards set at different levels restricts interpretation and comparison between studies and suggests that there is, as yet, no consensus on the level of language difficulty that needs to be identified by means of the screening process.
- It is uncertain whether the children identified by screening procedures of the type described here are the same children as those who stand to benefit from treatment. In general it seems likely, given the emphasis in the intervention section on expressive language delays, that the intervention is in most cases working for a subgroup of the children who would be identified by most screening procedures. Children identified by a screening procedure are likely to exhibit diverse speech and language skills with a wide range of social and psychological aetiologies.

Chapter 7
Conclusions

The evidence indicates that early speech and language delay should be a cause for concern to those involved with child health surveillance for the following reasons.

- The delay may pose problems for the individual child at the time of identification.
- The delay may indicate other co-morbid conditions such as hearing loss, developmental and behavioural difficulties.
- Early speech and language delay may have implications for the later development of literacy and socialisation.
- Evidence exists of the positive effects of intervention and, in particular, indirect treatment approaches that involve parents.

Given the current state of knowledge there is insufficient evidence to warrant the introduction of universal population screening for primary speech and language delay, due to the inherent problems associated with prevalence and natural history. However, as the reviewed data only considered universal population screening we cannot comment on the relative efficacy of alternative methods of early identification, which might include primary prevention, or confirmatory approaches driven by expressions of concern from parents and/or professionals. These alternative methods were beyond the limits of this review.

Implications for policy

There are many gaps in the literature, which make it impossible to make explicit service level recommendations. However, some suggestions have been formulated by the research team. (It should be noted that a number of studies were carried out in other countries with public health and educational services very different from those in the UK.)

The available evidence indicates that speech and language delays which persist into the school years remain an important problem for the individual, and for educational and health services. Although the available literature does not indicate that there is sufficient evidence to warrant the introduction of a screening programme, this does not mean that these children should not be identified. Rather, the responsibility for the identification might shift from the screen, where the burden of proof rests solely with the evidence supporting the screening measure, to a more mutual arrangement whereby parents consult primary-care professionals who, in turn, use their knowledge to assist in the identification process.

One method of investigating this approach would be to conceive of the early identification process in two stages. The first stage would be the elicitation of concern from the parent, and the second stage would be the application of an appropriately designed measure to reject or accept the initial concern. This approach would only work if parental judgement was sufficiently sensitive in the first instance. There remains some doubt whether this could be demonstrated, though it has yet to be properly investigated and is likely to be population-specific. The data from this review suggest that parent-focused measures may be as sensitive as screening tests. The priority for the screening procedure concerned would then be to maximise specificity. This possibility is supported by the present review, which suggests that the specificity of screening procedures tends to be both relatively high and generally more robust than sensitivity. Such a process would not necessarily meet the criteria for a screening procedure as envisaged by Wilson and Jungner (1968).

Despite these relatively robust findings for specificity, the number of children over-referred is likely to be high given the available prevalence estimates. This will prove costly to the services to whom these children are referred, and is likely to have negative implications for the child. In order to minimise the cost of over-referral to both the parent and the service provider it would be useful to explore the value of introducing experienced speech and language therapists as 'gate-keepers' to the speech and language therapy services. In practice, it is often the least experienced therapists who take on this role. Such a shift in emphasis would prove more costly (in that such staff are more expensive to employ), but it is likely to reduce the number of inappropriate referrals by exploiting the clinical experience of the individual speech and language therapists concerned.

The move away from formal screening procedures to a consultation between primary-care professional and parent would be likely to have implications for the training of health and educational professionals. In particular, primary-care professionals (health visitors, school nurses, nursery staff) would need to be made aware of the factors that help predict and mitigate against persistent problems, in order to identify children whose difficulties are least likely to resolve without intervention. This would involve a substantial training commitment, both in basic training and at the level of in-service training of the professionals concerned. The professional group likely to be most involved in the provision of such training would be speech and language therapists. There is currently little explicit recognition of the need for this type of support, particularly where clinical services are purchased according to individual contacts, and do not necessarily recognise the broader training commitment. This would need to be recognised by health commissioners.

Other forms of identification would also merit investigation. For example, the shift to the empowerment of parents may suggest that there may be a case for the distribution of appropriate information to all parents through, for example, the Personal Child Health Record, readily available literature, and other locally appropriate means of dissemination. Parents could then be encouraged to make use of 'drop-in' speech and language therapy clinics, which could then respond to parental need as it arises. While this may be an appealing option, the evidence in favour of it has yet to be examined.

There is also a case to be made for health visitors and other child-care professionals to be actively involved in the process of intervention with children where concern has been expressed. In some health trusts, health visitors are already involved in parent–child interaction programmes supporting a range of families in need. It would be appropriate to explore the possibility of increasing the remit of such groups to include speech and language work in the early years. This could be construed as a process of stopping children becoming cases, that is reducing incidence, rather than providing remediation, which seeks to reduce prevalence. This would constitute primary rather than secondary prevention.

At this stage, despite some promising data from early intervention studies, there is insufficient evidence for prioritising very early intervention (under 2 years) relative to later intervention for primary speech and language delays. Effectively remediating more entrenched difficulties in the early school years using intervention approaches of confirmed validity may be a more realistic use of resources than concentrating on very early delays with an uncertain prognosis. Progress is likely to be made on this issue once Ward's follow-up study of children identified and treated in the second year of life has been completed.

The data support the adoption of indirect models of intervention relative to the more traditional direct models for all language difficulties, though not for speech delays. There needs to be further exploration of the effectiveness of different intermediaries, such as parents, teachers and nursery staff.

The wide adoption of indirect approaches to intervention would have considerable implications for the speech and language therapy profession, adding momentum to the shift to the consultative role for the speech and language therapist. This, in turn, is likely to lead to modification of the basic training for speech and language therapists. Increased emphasis would need to be placed on teaching related to adult learning styles to reflect the shift of the focus of intervention from the child to the intermediary.

Indirect methods are likely to require increased levels of active involvement in the therapeutic process on the part of the intermediary. This may prove unduly onerous for some people who may be unwilling to accept such a commitment. This will almost certainly raise the question of how the service best meets the needs of the children concerned.

Recommendations for research

While it is possible to address some of the research issues within a specific educational or medical paradigm, the authors would wish to stress the need for interagency collaboration at a research level.

Studies

There is a need for a systematic review to provide complementary data on educational outcomes. This would be of considerable value to educational and health services and the joint funding of such a review would provide a useful starting point for the development of a combined research strategy between the two agencies.

There is a good case for a systematic review of evidence comparing interventions, including treatment 1 – treatment 2 – no control designs,

which were not included in the present study. Such a review would prove very valuable in informing clinical and educational practice relating to speech and language delay.

Prevalence

Prevalence rates need to be established in different populations using an agreed definition of case status. It is likely that such a definition will involve a composite of performance on standardised procedures and clinical judgement. It will be particularly important to establish differential levels of need across social classes and in bilingual and ethnically diverse populations. As a first step, there is a need for the collection of national figures with a view to developing a consensus of which children need to be treated.

Natural history

The issue of natural history is inherently problematic because it is almost impossible to factor out the effects of environmental modifications, be they educational or therapeutic. There is a need to develop predictive models which incorporate levels of intervention together with linguistic, familial and neurodevelopmental information.

Intervention

Controlled trials and complementary experimental single-subject studies are required. These should address the following issues:

- the extent to which it is possible to provide successful interventions for children whose difficulties would not otherwise resolve spontaneously
- the relative value of early versus late intervention
- the effects of the same intervention approach on different combinations of symptoms (i.e. receptive/expressive delays versus expressive delays only)
- the application of different approaches to the same combination of symptoms.
- the differential effects for different sub-groups in the population.

This review raises the possibility of synthesising intervention outcomes to develop 'benchmarks' of therapeutic effect. Further work could explore the value of developing such indicators as performance measures in clinical audit.

Given the potential value of indirect intervention, it is important that studies are carried out identifying the characteristics of parents and carers who are most likely to succeed with indirect intervention approaches. It is also important to evaluate alternative methods of working with potential intermediaries who cannot make such a commitment to therapy.

The value of a shift from secondary to primary prevention in the very early years would need further investigation before any recommendation regarding such an approach could be made.

Future studies need to be designed with sufficient sample size to ensure adequate statistical power and provide protection against Type II error.

Screening

An identification procedure with good predictive validity would make a considerable contribution to this field. This may involve providing better data on existing measures. Alternatively, there is a case for incorporating findings from natural history and intervention studies into a scale that could be used by primary-care workers and parents.

The identification process itself needs to be examined as a comparison between procedures applied to single populations and as replications across different populations using the same parameters. Where possible, such studies need to be carried out in the context of intervention studies.

Given that screening for early speech and language delay does not meet the criteria required for the introduction of a universal screening programme but that there is a need to identify the cases concerned, it is necessary to investigate alternative methods of identification. In particular it would be appropriate to consider identification procedures that place more emphasis on the clinical judgement of the primary-care worker and on the role of the carer. As an extension of this it would be valuable to examine the effectiveness of a two-stage approach to identification, the first stage involving informed parental judgement, the second a confirmatory screening procedure or an explicit diagnosis by an experienced clinician.

As part of the investigation of the screening process, particular attention needs to be paid to the relative costs and utilities of different screen outcomes. In particular, attention needs to be paid to false-positive and false-negative outcomes and the effect that this has on the child, on the family and on the services. As a preliminary stage, more data need to be collected on current service structures and their costs.

Once a robust measure with good predictive validity has been developed it will be appropriate to carry out a screen/no-screen trial.

Acknowledgements

Guidance in the systematic review process from Professor Trevor Sheldon, Director, NHS Centre for Reviews and Dissemination, York University, and from The Cochrane Centre, Oxford is gratefully acknowledged.

The review team were assisted by the many valuable comments of the Advisory Group: Ms Pauline Beirne, Speech and Language Therapy Manager; Ms Norma Corkish, Director, Association for All Speech Impaired Children; Professor Cliff Cunningham, Psychologist; Dr Elizabeth Dean, Speech and Language Therapist; Professor Pam Enderby, Speech and Language Therapist; Professor David Hall, Paediatrician; Dr Stuart Logan, Epidemiologist; Dr Kay Mogford-Bevan, Speech and Language Therapist/Psychologist; Professor Colin Robson, Psychologist; Dr Sally Ward, Speech and Language Therapist.

Librarians who enabled the review: Ms M Plackett, RNID Library, University College London; Mr Friend, Librarian, University College London; Ms G Morris and Ms M McNulty, Librarians, University of Strathclyde; Ms I Stirling, Information Services Librarian, Jordanhill Library, University of Strathclyde; Ms C Muller, Librarian, City University.

Those giving advice on literature searches and on access to database resources: Mr C Norris, Biomedical Research Indexing; Ms Jones, NHS Centre for Reviews and Dissemination, York University; Ms Johnson, Research Librarian, King's Fund, for access to Cochrane Library databases; Ms M Grant, Information Officer, UK Clearing House on Health Outcomes, Leeds.

Those giving advice on meta-analysis techniques: (regarding screening) Professor D Sackett NHS R&D Centre for Evidence-based Medicine, Oxford; Professor L Irwig, Associate Professor (Epidemiology), University of Sydney, Australia.

Those who supplied literature or bibliographies of particular interest to the review: Professor Dr PH Dejonckere, Academisch Ziekenhuis Utrecht, The Netherlands; Dr F Glascoe, Associate Professor of Pediatrics, Vanderbilt University, USA; Dr H Ireton, Behavior Science Systems Inc., Minneapolis, USA; Professor T Marteau, Director and Professor of Health Psychology, UMDS, Guy's and St Thomas's Medical and Dental School, London, UK; Dr RA Sturner, Associate Professor of Pediatrics, John Hopkins University, USA.

Technical support was provided by Mr T Gurney, Computer Officer, Department of Psychology, University of Strathclyde.

Statistical support was provided by Dr W Cheyne, Department of Psychology, University of Strathclyde.

Thanks go to the authors who kindly responded to our requests for further information on their data, and those who allowed use of their unpublished manuscripts.

References

Studies included in the review

A. Prevalence

Bax M, Hart H, Jenkins S. The behaviour, development and health of the young child: implications for care. *BMJ* 1983;**286**:1793–6.

Bax M, Hart H, Jenkins S. The health needs of the pre-school child. 1980 (Unpublished).

Beitchman JH, Nair R, Clegg M, Patel PG, Ferguson B, Pressman E, *et al.* Prevalence of speech and language disorders in 5-year-old kindergarten children in the Ottawa-Carleton region [published erratum appears in *J Speech Hear Disord* 1987;**52**(1):94]. *J Speech Hear Disord* 1986;**51**(2):98–110.

Burden V, Stott CM, Forge J, Goodyer I. The Cambridge Language and Speech Project (CLASP). 1. Detection of language difficulties at 36–39 months. *Dev Med Child Neurol* 1996;**38**(7):613–31.

Dudley JG, Delage J. Incidence des troubles de la parole et du langage chez les enfants franco-quebecois. *Communication Humaine* 1980;**5**:131–42.

Harasty J, Reed VA. The prevalence of speech and language impairment in two Sydney metropolitan schools. *Aust J Hum Commun Disord* 1994;**22**(1):1–23.

Kirkpatrick E, Ward J. Prevalence of articulation errors in New South Wales primary school pupils. *Aust J Hum Commun Disord* 1984;**12**:55–62.

Paul TJ, Desai P, Thorburn MJ. The prevalence of childhood disability and related medical diagnoses in Clarendon, Jamaica. *West Indian Med J* 1992;**41**(1):8–11.

Randall D, Reynell J, Curwen M. A study of language development in a sample of 3 year old children. *Br J Disord Commun* 1974;**9**(1):3–16.

Rescorla L, Hadicke-Wiley M, Escarce E. Epidemiological investigation of expressive language delay at age two. Special Issue: Language development in special populations. *First Lang* 1993;**13**(37 Pt 1):5–22.

Richman N, Stevenson J, Graham PJ. Preschool to school: a behavioural study. London: Academic Press, 1982.

Silva PA. The prevalence, stability and significance of developmental language delay in preschool children. *Dev Med Child Neurol* 1980;**22**:786–77.

Silva PA, McGee R, Williams SM. Developmental language delay from three to seven years and its significance for low intelligence and reading difficulties at age seven. *Dev Med Child Neurol* 1983;**25**:783–93.

Stevenson J, Richman N. The prevalence of language delay in a population of three-year-old children and its association with general retardation. *Dev Med Child Neurol* 1976;**18**(4):431–41.

Stewart JM, Hester EJ, Taylor DL. Prevalence of language, speech and hearing disorders in an urban preschool black population. *J Child Commun Disord* 1986;**9**(2):107–23.

Thorburn MJ, Desai P, Durkin M. A comparison of efficacy of key informant and community survey methods in the identification of childhood disability in Jamaica. *Ann Epidemiol* 1991;**1**:255–61.

Tomblin JB, Smith E, Zhang X. Epidemiology of specific language impairment: prenatal and perinatal risk factors. *J Commun Disord* 1997;**30**:325–440.

Tomblin JB, Records NL, Zhang X. A system for the diagnosis of specific language impairment in kindergarten children. *J Speech Hear Res* 1996;**39**(6):1284–94.

Tuomi S, Ivanoff P. Incidence of speech and hearing disorders among kindergarten and grade 1 children. *Special Education in Canada* 1977;**51**(4):5–8.

Warr-Leeper GA, McShea RS, Leeper HA, Jr. The incidence of voice and speech deviations in a middle school population. *Lang Speech Hear Serv Schools* 1979;**10**(1):14–20.

Wong V, Lee PWH, Mak-Lieh F, Yeung CY, Leung PWL, Luk SL, *et al.* Language screening in preschool Chinese children. *Eur J Disord Commun* 1992;**27**(3):247–64.

B. Natural history

Bralley RC, Stoudt RJ. A five year longitudinal study of development of articulation. *Lang Speech Hear Serv Schools* 1977;**8**(3):176–80.

Felsenfeld S, Broen PA, McGue M. A 28-year follow-up of adults with a history of moderate phonological disorder: linguistic and personality results. *J Speech Hear Res* 1992;**35**:1114–25.

Fiedler MF, Lenneberg EH, Rolfe UT, Drorbaugh JE. A speech screening procedure with three-year-old children. *Pediatrics* 1971;**48**(2):268–76.

Hall NE. Language and fluency in child language disorders: changes over time. *J Fluency Disord* 1996;**21**:1–32.

Hall NE, Yamashita TS, Aram DM. Relationship between language and fluency in children with developmental language disorders. *J Speech Hear Res* 1993;**36**:568–79.

Klee T, Carson DK, Gavin WJ, Hall L, Kent A, Reece S. Concurrent and predictive validity of an early language screening program. 1997 (Unpublished).

Lancaster G. The effectiveness of parent administered input training for children with phonological training for children with phonological disorders [MSc thesis]. London (UK): City University, 1991.

Renfrew CE, Geary L. Prediction of persisting speech deficit. *Br J Disord Commun* 1973;**8**(1):37–41.

Rescorla L, Schwartz E. Outcome of toddlers with specific expressive language delay. *Appl Psycholinguist* 1990;**11**(4):393–407.

Richman N, Stevenson J, Graham PJ. Preschool to school: a behavioural study. London: Academic Press, 1982.

Scarborough HS, Dobrich W. Development of children with early language delay. *J Speech Hear Res* 1990;**33**(1):70–83.

Silva PA. The prevalence, stability and significance of developmental language delay in preschool children. *Dev Med Child Neurol* 1980;**22**:768–77.

Silva PA, McGee R, Williams SM. Developmental language delay from three to seven years and its significance for low intelligence and reading difficulties at age seven. *Dev Med Child Neurol* 1983;**25**:783–93.

Silva PA, Williams SM, McGee R. A longitudinal study of children with developmental language delay at age three: later intelligence, reading and behaviour problems. *Dev Med Child Neurol* 1987;**29**:630–40.

Stackhouse J, Snowling M. Developmental verbal dyspraxia II: a developmental perspective on two case studies. *Eur J Disord Commun* 1992;**27**(1):35–54.

Stevenson J, Richman N. The prevalence of language delay in a population of three-year-old children and its association with general retardation. *Dev Med Child Neurol* 1976;**18**(4):431–41.

Stevenson P, Bax M, Stevenson J. The evaluation of home-based speech therapy for language delayed preschool children in an inner city area. *Br J Disord Commun* 1982;**17**(3):141–8.

Thal DJ, Bates E. Language and gesture in late talkers. *J Speech Hear Res* 1988;**31**:115–23.

Thal DJ, Tobias S. Communicative gestures in children with delayed onset of oral expressive vocabulary. *J Speech Hear Res* 1992;**35**:1281–9.

Thal DJ, Tobias S, Morrison D. Language and gesture in late talkers: a one year follow-up. *J Speech Hear Res* 1991;**34**:604–12.

Ward, S. Validation of a treatment method. 2nd Conference of the Comite Permanente de Liaison des Othophonistes-Logopedes del'UE [CPLOL], Antwerp 1994.

Ward S. The predictive validity and accuracy of a screening test for language delay and auditory perceptual disorder. *Br J Disord Commun* 1992;**27**:55–72.

C. Intervention

Almost D, Rosenbaum P. Effectiveness of speech intervention for phonological disorders: a randomized controlled trial. *Dev Med Child Neurol* 1998;**40**:319–52.

Bedrosian JL, Willis TL. Effects of treatment on the topic performance of a school-age child. *Lang Speech Hear Serv Schools* 1987;**18**(2):158–67.

Camarata S. The application of naturalistic conversation training to speech production in children with speech disabilities. *J Appl Behav Anal* 1993;**26**(2):173–82.

Conant S, Budoff M, Hecht B, Morse R. Language intervention: a pragmatic approach. *J Autism Dev Disord* 1984;**14**(3):301–17.

Connell PJ. Acquisition of semantic role by language-disordered children: difference between production and comprehension. *J Speech Hear Res* 1986a;**29**:366–74.

Connell PJ. Teaching subjecthood to language-disordered children. *J Speech Hear Res* 1986b;**29**:481–92.

Culatta B, Horn D. A program for achieving generalization of grammatical rules to spontaneous discourse. *J Speech Hear Disord* 1982;**47**(2):174–80.

Dollaghan C, Kaston N. A comprehension monitoring program for language-impaired children. *J Speech Hear Disord* 1986;**51**(3):264–71.

Ellis Weismer S, Murray-Branch J. Modelling versus modelling plus evoked production training: a comparison of two language intervention methods. *J Speech Hear Disord* 1989;**54**(2):269–81.

Ellis Weismer S, Murray-Branch J, Miller JF. Comparison of two methods for promoting productive vocabulary in late talkers. *J Speech Hear Res* 1993;**36**:1037–50.

Fey ME, Cleave PL, Long SH, Hughes DL. Two approaches to the facilitation of grammar in children with language impairment: an experimental evaluation. *J Speech Hear Res* 1993;**36**:141–57.

Fey ME, Cleave PL, Ravida AI, Long SH, Dejmal AE, Easton DL. Effects of grammar facilitation on the phonological performance of children with speech and language impairments. *J Speech Hear Res* 1994;**37**:594–607.

Gibbard D. Parental-based intervention with pre-school language-delayed children. *Eur J Disord Commun* 1994;**29**(2):131–50.

Gierut JA. Differential learning of phonological oppositions. *J Speech Hear Res* 1990;**33**(3):540–9.

Gierut JA. The conditions and course of clinically induced phonological change. *J Speech Hear Res* 1992;**35**(5):1049–63.

Girolametto L, Pearce PS, Weitzman E. Interactive focused stimulation for toddlers with expressive vocabulary delays. *J Speech Hear Res* 1996;**39**(6):1274–83.

Girolametto L, Pearce PS, Weitzman E. The effects of focused stimulation for promoting vocabulary in young children with delays: a pilot study. *J Child Commun Dev* 1995;**17**(2):39–49.

Hargrove PM, Roetzel K, Hoodin RB. Modifying the prosody of a language-impaired child. *Lang Speech Hear Serv Schools* 1989;**20**:245–58.

Hegde MN, Gierut J. The operant training and generalization of pronouns and a verb form in a language delayed child. *J Commun Disord* 1979;**12**(1):23–34.

Hegde MN, Noll MJ, Pecora R. A study of some factors affecting generalization of language training. *J Speech Hear Disord* 1979;**44**(3):301–20.

Hemmeter ML, Kaiser AP. Enhanced milieu teaching: effects of parent-implemented language intervention. *J Early Intervention* 1994;**18**(3):269–89.

Kaiser AP, Hemmeter ML, Ostrosky MM, Alpert CL, Hancock TB. The effects of group training and individual feedback on parents use of milieu teaching. *J Child Commun Disord* 1995;**16**(2):39–48.

Kaiser AP, Hester PP, Alpert CL, Whiteman BC. Preparing parent trainers: an experimental analysis of effects on trainers, parents, and children. *Topics in Early Childhood Special Education* 1995;**15**(4):385–414.

Lancaster G. The effectiveness of parent administered input training for children with phonological disorders [MSc thesis]. London (UK): City University, 1991.

Matheny N, Panagos JM. Comparing the effects of articulation and syntax programs on syntax and articulation improvement. *Lang Speech Hear Serv Schools* 1978;**9**:57–61.

McDade A, McCartan PA. 'Partnership with parents'. A pilot study. 1996 (Unpublished).

McGregor KK. Use of phonological information in a word-finding treatment for children. *J Speech Hear Res* 1994;**37**(6):1381–93.

Olswang L, Bain B, Dunn C, Cooper J. The effects of stimulus variation on lexical learning. *J Speech Hear Disord* 1983;**48**:192–201.

Olswang LB, Bain BA. Monitoring phoneme acquisition for making treatment withdrawal decisions. *Appl Psycholinguist* 1985;**6**(1):17–37.

Olswang LB, Bain BA, Rosendahl PD, Oblak SB, Smith AE. Language learning: moving performance from a context-dependent to context-independent state. *Child Lang Teaching Ther* 1986;**2**(2):180–210.

Olswang LB, Coggins TE. The effects of adult behaviours on increasing language delayed children's production of early relational meanings. *Br J Disord Commun* 1984;**19**:15–34.

Powell TW, Elbert M. Generalization following the remediation of early- and later-developing consonant clusters. *J Speech Hear Disord* 1984;**49**(2):211–18.

Powell TW, Elbert M, Dinnsen DA. Stimulability as a factor in the phonological generalization of misarticulating preschool children. *J Speech Hear Res* 1991;**34**(6):1318–28.

Reid J, Donaldson ML, Howell J, Dean EC, Grieve R. The effectiveness of therapy for child phonological disorder: The Metaphon Approach. In: Aldridge M, editor. Child Language. Clevedon, Avon: Multilingual Matters, 1996:165–75.

Schwartz RG, Chapman K, Terrell BY, Prelock P, Rowan L. Facilitating word combination in language-impaired children through discourse structure. *J Speech Hear Disord* 1985;**50**:31–9.

Shelton RL, Johnson AF, Ruscello DM, Arndt WB. Assessment of parent-administered listening training for preschool children with articulation deficits. *J Speech Hear Disord* 1978;**43**(2):242–54.

Stevenson P, Bax M, Stevenson J. The evaluation of home-based speech therapy for language delayed preschool children in an inner city area. *Br J Disord Commun* 1982;**17**(3):141–8.

Ward, S. Validation of a treatment method. 2nd Conference of the Comite Permanente de Liaison des Othophonistes-Logopedes del'UE [CPLOL], Antwerp 1994.

Warren SF, McQuarter RJ, Rogers-Warren A. The effects of mands and models on the speech of unresponsive language-delayed preschool children. *J Speech Hear Disord* 1984;**49**(1):43–52.

Warrick N, Rubin H, Rowe-Walsh S. Phoneme awareness in language-delayed children: comparative studies and intervention. *Ann Dyslexia* 1993;**43**:153–73.

Whitehurst GJ, Fischel JE, Lonigan CJ, Valdez-Menchaca MC, Arnold DS, Smith M. Treatment or early expressive language delay: if, when, and how. *Topics in Language Disorders* 1991;**11**(4):55–68.

Wilcox MJ, Leonard LB. Experimental acquisition of Wh- questions in language-disordered children. *J Speech Hear Res* 1978;**21**:220–39.

Williams AL. Generalization patterns associated with training least phonological knowledge. *J Speech Hear Res* 1991;**34**(4):722–33.

Young EC. The effects of treatment on consonant cluster and weak syllable reduction processes in misarticulating children. *Lang Speech Hear Serv Schools* 1987;**18**(1):23–33.

Zwitman DH, Sonderman JC. A syntax program designed to present base linguistic structures to language-disordered children. *J Commun Disord* 1979;**12**(4):323–35.

D. Screening

AFASIC. The AFASIC checklists: User's Guide. Cambridge, UK: LDA, 1991.

Allen DV, Bliss LS. Concurrent validity of two screening tests. *J Commun Disord* 1987;**20**(4):305–17.

Black MM, Gerson LF, Freeland CAB, Nair P, Rubin JS, Hutcheson JJ. Language screening for infants prone to otitis media. *J Pediatr Psychol* 1988;**13**(3):423–33.

Blaxley L, Clinker M, Warr-Leeper G. Two language screening tests compared with developmental sentence scoring. *Lang Speech Hear Serv Schools* 1983;**14**:38–46.

Bliss LS, Allen DV. Screening Kit of Language Development: a preschool language screening instrument. *J Commun Disord* 1984;**17**(2):133–41.

Borowitz KC, Glascoe FP. Sensitivity of the Denver Developmental Screening Test in speech and language screening [see letters]. *Pediatrics* 1986;**78**(6):1075–78.

Chaffee CA, Cunningham CE. Screening effectiveness of the Minnesota Child Development Inventory Expressive and Receptive Language Scales: sensitivity, specificity, and predictive value. *Psychol Assess* 1990;**2**(1):80–5.

Chevrie-Muller C, Simon AM, Dufouil C, Goujard J. [Early screening of language acquisition disorders in the 3.5-year-old: validation of a method]. *Approche Neuropsychologique des Apprentissages chez l'Enfant* 1993;**5**(2):82–91.

Clark JG, Jorgensen SK, Blondeau R. Investigating the validity of the Clinical Linguistic Auditory Milestone Scale. *Int J Pediatr Otorhinolaryngol* 1995;**31**(1):63–75.

Coplan J, Gleason JR, Ryan R, Burke MG, Williams ML. Validation of an Early Language Milestone Scale in a high-risk population. *Pediatrics* 1982;**70**(5):677–83.

Culatta B, Page JL, Ellis J. Story retelling as a communicative performance screening tool. *Lang Speech Hear Serv Schools* 1983;**14**:66–74 [Comment in *Lang Speech Hear Serv Schools* 1984;**15**:140–1].

Dale PS, Henderson VL. An evaluation of the Test of Early Language Development as a measure of receptive and expressive language. *Lang Speech Hear Serv Schools* 1987;**18**(2):179–87.

Dixon J, Kot A, Law J. Early language screening in City and Hackney: work in progress. *Child Care Health Dev* 1988;**14**(3):213–29.

Dodge GR. A comparison of language screening methods. *Lang Speech Hear Serv Schools* 1980;**11**:214–17.

Downs M, Blager FB. The otitis prone child. *J Dev Behav Pediatr* 1982;**3**(2):106–13.

Feeney J, Bernthal J. The efficiency of the revised Denver Developmental Screening Test as a language screening tool. *Lang Speech Hear Serv Schools* 1996;**27**(4):330–32.

German ML, Williams E, Herzfeld J, Marshall RM. Utility of the Revised Denver Developmental Screening Test and the Developmental Profile II in identifying preschool children with cognitive, language, and motor problems. *Education and Training of the Mentally Retarded* 1982;**17**(4):319–24.

Glascoe FP. Can clinical judgment detect children with speech-language problems? *Pediatrics* 1991;**87**(3):317–22 [Comment in *Pediatrics* 1991;**88**(4):875–6].

Glascoe FP, Byrne KE. The accuracy of three developmental screening tests. *J Early Intervention* 1993;**17**(4):368–79.

Haber JS, Norris M. The Texas Preschool Screening Inventory: a simple screening device for language and learning disorders. *Child Health Care* 1983;**12**(1):11–18.

Illerbrun D, Haines L, Greenough P. Language Identification Screening Test for Kindergarten: a comparison with four screening and three diagnostic language tests. *Lang Speech Hear Serv Schools* 1985;**16**(4):280–91.

Klee T, Carson D, Kent A, Reece S. Concurrent and predictive validity of an early language screening program. Symposium on Research in Child Language Disorders; 1996; Madison.

Klee T, Carson DK, Gavin WJ, Hall L, Kent A, Reece S. Concurrent and predictive validity of an early language screening program. 1997 (Unpublished).

Law J. Early language screening in City and Hackney: the concurrent validity of a measure designed for use with 2½-year-olds. *Child Care Health Dev* 1994;**20**(5):295–308.

Law J. Two and a half year olds in Hackney, London: the early identification of language impairment and a description of related characteristics [PhD Thesis]. London, UK: City University, 1993.

Levett L, Muir J. Which three year olds need speech therapy? Uses of the Levett-Muir language screening test. *Health Visitor* 1983;**56**:454–6.

McGinty C. An investigation into aspects of the Mayo Early Language Screening Test [MSc Thesis]. Dublin, EIRE: Dublin University, 1996.

Rescorla L. The Language Development Survey: a screening tool for delayed language in toddlers. *J Speech Hear Disord* 1989;**54**(4):587–99.

Rescorla L, Hadicke-Wiley M, Escarce E. Epidemiological investigation of expressive language delay at age two. Special Issue: Language development in special populations. *First Lang* 1993;**13**(37 Pt 1):5–22.

Richman N, Stevenson J, Graham PJ. Preschool to school: a behavioural study. London: Academic Press, 1982.

Rigby MJ, Chesham I. A trial speech screening test for school entrants. *BMJ* 1981;**282**:449–51.

Scherer NJ, D'Antonio LL. Parent questionnaire for screening early language development in children with cleft-palate. *Cleft Palate Craniofacial J* 1995;**32**(1):7–13.

Stevenson J, Richman N. The prevalence of language delay in a population of three-year-old children and its association with general retardation. *Dev Med Child Neurol* 1976;**18**(4):431–41.

Stokes SF. Secondary prevention of paediatric language disability: a comparison of parents and nurses. *Eur J Disord Commun* 1997;**32**(2):139–59.

Sturner RA, Funk SG, Green JA. Preschool speech and language screening: further validation of the Sentence Repetition Screening Test. *J Dev Behav Pediatr* 1996;**17**(6):405–13.

Sturner RA, Heller JH, Funk SG, Layton TL. The Fluharty Preschool Speech and Language Screening Test: a population-based validation study using sample-independent decision rules. *J Speech Hear Res* 1993a;**36**(4):738–45.

Sturner RA, Kunze L, Funk SG, Green JA. Elicited imitation: its effectiveness for speech and language screening. *Dev Med Child Neurol* 1993b;**35**(8):715–26.

Walker D, Gugenheim S, Downs M, Northern JL. Early Language Milestone Scale and language screening of young children. *Pediatrics* 1989;**83**(2):284–8 [Comment in *Pediatrics* 1989;**84**(3):586–7].

Ward S. The predictive validity and accuracy of a screening test for language delay and auditory perceptual disorder. *Eur J Disord Commun* 1992;**27**(1):55–72.

Ward S. Detecting abnormal auditory behaviours in infancy: the relationship between such behaviours and linguistic development. *Br J Disord Commun* 1984;**19**(3):237–51.

Ward S. An investigation into the early detection and diagnosis of congenital auditory imperception [PhD Thesis]. Birmingham, UK: City of Birmingham Polytechnic, 1986.

Ward S, Birkett D. The Ward Infant Language Screening Test, Assessment, Acceleration and Remediation. Manchester, UK: Central Manchester Health Care Trust, 1994.

Westerlund M, Frylmark A. Evaluation of a screening programme. Conference of the Comite Permanente de Liaison des Othophonistes-Logopedes del'UE [CPLOL], Lisbon, 1997.

Westerlund M. Screening for speech and language in three-year-olds: experiences from a field study in a Swedish municipality. 1996 (Unpublished).

Whitworth A, Daves C, Stokes S, Blain T. Identification of communication impairments in preschoolers: a comparison of parent and teachers success. *Aust J Hum Commun Disord* 1993;**21**(1):112–33.

Studies excluded from the review

A. Prevalence

Aithal V. Prevalence of speech, language and hearing disorders. *Hear Aid J* 1985;**4**(3):68–73. Exclude: 5 6.

Bendel J, Palti H, Winter S, Ornoy A. Prevalence of disabilities in a national sample of 3-year-old Israeli children. *Isr J Med Sci* 1989;**25**(5):264–70. Exclude: 4 6.

Boyle CA, Decoufle P, Yeargin-Allsopp M. Prevalence and health impact of developmental disabilities in US children. *Pediatrics* 1994;**93**(3):399–403. Exclude: 4.

Burd L, Hammes K, Bornhoeft DM, Fisher W. A North Dakota prevalence study of nonverbal school-age children. *Lang Speech Hear Serv Schools* 1988;**19**(4):371–83. Exclude: 4 6.

Butler NR, Peckham C, Sheridan M. Speech defects in children aged 7 years: a national study. *BMJ* 1973;**3**:253–7. Exclude: 4.

Calnan M, Richardson K. Speech problems in a national survey: assessments and prevalences. *Child Care Health Dev* 1976;**2**(4):181–202. Exclude: 4.

Canning PM, Lyon ME. Young children with special needs: prevalence and implications in Nova Scotia. *Can J Ed* 1989;**14**(3):368–80. Exclude: 4 6.

Eu B. Health of the preschool child. *Med J Aust* 1979;**2**:84–6. Exclude: 4 5 6.

Fundudis T, Kolvin I, Garside RF, editors. Speech retarded and deaf children: their psychological development. London: Academic Press, 1979. Exclude: 4 6.

Gillespie SK, Cooper EB. Prevalence of speech problems in junior and senior high schools. *J Speech Hear Res* 1973;**16**(4):739–43. Exclude: 4.

Grinnell SW, Scott-Hartnett D, Glasier JL. Language disorders. *J Am Acad Child Psychiatry* 1983;**22**(6):580–1. Exclude: 5.

Haimanot RT, Abebe M, Mariam AG, Forsgren L, Holmgren G, Heijbel J, *et al.* Community-based study of neurological disorders in Ethiopia: development of a screening instrument. *Ethiop Med J* 1990;**28**(3):123–37. Exclude: 4 6.

Haller RM, Thompson EA. Prevalence of speech, language, and hearing disorders among Harlem children. *J Nat Med Assoc* 1975;**67**(4):298–301, 325. Exclude: 4 5 6.

Hull FM, Mielke PW, Timmons RJ, Willeford JA. The national speech and hearing survey. *ASHA* 1971;**13**:501–9. Exclude: 4 6.

Jackson AL. Prevalence of selected impairments. *Vital Health Stat [1]* 1968;**10**(48):1–78. Exclude: 4 6.

Johnston O. Ill health and developmental delays in Adelaide four-year-old children. *Aust Paediatr J* 1980;**16**(4):248–54. Exclude: 4 6.

Lehingue Y, Fassio F, Momas I, Daures JP. Etat de sante des enfants d'ecole maternelle. [Epidemiological surveillance of nursery school children in the Herault for the evaluation of maternal and infant health and welfare services]. *Revue d'Epidemiologie et de Sante Publique* 1992;**40**(1):25–32. Exclude: 6.

Luotonen M. Early speech development, articulation and reading ability up to the age of 9. *Folia Phoniatrica et Logopaedica* 1995;**47**(6):310–17. Exclude: 4.

Mathur GP, Mathur S, Singh YD, Kushwaha KP, Lele SN. Detection and prevention of childhood disability with the help of Anganwadi workers. *Indian Pediatr* 1995;**32**(7):773–7. Exclude: 4 6.

Miller L, Kiernan MT, Mathers MI, Klein-Gitelman M. Developmental and nutritional status of internationally adopted children. *Arch Pediatr Adolesc Med* 1995;**149**(1):40–4. Exclude: 5.

Nowotny M, Stretton PJ. The health of the preschool child. *Med J Aust* 1979;**1**:289–91. Exclude: 4 6.

Palti H, Bendel J, Ornoy A. Prevalence of disabilities in a national sample of 7-year-old Israeli children. *Isr J Med Sci* 1992;**28**(7):435–41. Exclude: 4 6.

Peckham CS. Speech defects in a national sample of children aged seven years. *Br J Disord Commun* 1973;**8**(1):2–8. Exclude: 4.

Ratto A, Diaco P, Perillo L. Indagine epidemiologia sulle disfonie in eta scolare. Data preliminari. [Epidemiological survey of dysphonia in school children. Preliminary findings]. *Arch Stomatol Napoli* 1990;**31**(1):85–91. Exclude: 4 6.

Senecal J, Roussey M, Letort F, Delahaye M, Subileau B. Les causes d'inadaptation scolaire chez l'enfant. Depistage en maternelle et evolution [Screening for school difficulties in kindergarten children]. *Archives Francaises de Pediatrie* 1987;**44**(7):523–30. Exclude: 4 6.

Solomons G. Prevalence of speech and hearing problems in a child development clinic population. *Clin Pediatr (Phila)* 1970;**9**(7):384–9. Exclude: 5.

Stewart JM, Martin ME, Brady GM. Communicative disorders at a health-care center. *J Commun Disord* 1979;**12**(4):349–59. Exclude: 5.

Stewart JM, Spells VR. Current trends of communicative disorders in a public school system. *Aust J Hum Commun Disord* 1982;**10**:33–57. Exclude: 4 6.

Sugawara A. Fact-finding of voice and speech disorders in school children in Yokosuka City, Japan. 16th International congress of Logopedics and Phoniatrics; 1976; Interlaken. Exclude: 4 6.

Sundelin C, Vuille JC. Health screening of four year olds in a Swedish county: I. Organisation, methods, and participation. *Acta Paediatr Scand* 1975;**64**:795–800. Exclude: 4 6.

Sundelin C, Vuille JC. Health screening of four year olds in a Swedish county: II. Effectiveness in detecting health problems. *Acta Paediatr Scand* 1975;**64**:801–6. Exclude: 4 6.

Williams DM, Darbyshire JO, Vaghy DA. An epidemiological study of speech and hearing disorders. *J Otolaryngol Suppl* 1980;**7**:1–24. Exclude: 6.

B. Natural history

Fischel JE, Whitehurst GJ, Caulfield MB, DeBaryshe B. Language growth in children with expressive language delay. *Pediatrics* 1989;**82**(2):218–27. Exclude: 6.

Fundudis T, Kolvin I, Garside RF, editors. Speech retarded and deaf children:their psychological development. London: Academic Press, 1979. Exclude: 7.

Levi G, Capozzi F, Fabrizi A, Sechi E. Language disorders and prognosis for reading disabilities in developmental age. *Percept Mot Skills* 1982;**54**:1119–22. Exclude: 5 7.

Randall D, Reynell J, Curwen M. A study of language development in a sample of three year old children. *Br J Disord Commun* 1974;**9**(1):3–16. Exclude: 7.

Sax MR. A longitudinal study of articulation change. *Lang Speech Hear Serv Schools* 1972;**3**:41–8. Exclude: 7.

C. Intervention

Alpert CL, Kaiser AP. Training parents as milieu language teachers. *J Early Intervention* 1992;**16**(1):1–52. Exclude: 7.

Anderson C. Phonological speech problems: extending the applications for Metaphon therapy. 1997, (Unpublished). Exclude: 7.

Barratt J, Littlejohns P, Thompson J. Trial of intensive compared with weekly speech therapy in preschool children. *Arch Dis Child* 1992;**67**:106–8. Exclude: 7.

Barbeito CL. A comparison of two methods for teaching prepositions to language handicapped preschool children [PhD thesis]. Denver: University of Denver, 1972. Exclude: 7.

Bernhardt B. The application of nonlinear phonological theory to intervention with one phonologically disordered child. *Clin Linguist Phonetics* 1992;**6**(4):283–316. Exclude: 7.

Bobkoff K, Panagos JM. The 'point' of language intervention lessons. *Child Lang Teaching Ther* 1986;**2**(1):50–62. Exclude: 7.

Bouillon KR. The comparative efficacy of non-directive group play therapy with preschool speech or language delayed children [PhD thesis]. Texas Tech University, 1973. Exclude: 6.

Brown BB. Language vulnerability, speech delay and therapeutic intervention. *Br J Disord Commun* 1976;**11**(1):43–56. Exclude: 7.

Bryan A, Howard D. Frozen phonology thawed: the analysis and remediation of a developmental disorder of real word phonology. *Eur J Disord Commun* 1992;**27**(4):343–65. Exclude: 7.

Bunce BH, Ruder KF, Ruder CC. Using the miniature linguistic system in teaching syntax: two case studies. *J Speech Hear Disord* 1985;**50**(3):247–53. Exclude: 7.

Camarata SM, Nelson KE. Treatment efficiency as a function of target selection in the remediation of child language disorders. *Clin Linguist Phonetics* 1992;**6**(3):167–78. Exclude: 7.

Camarata SM, Nelson KE, Camarata MN. Comparison of conversational-recasting and imitative procedures for training grammatical structures in children with specific language impairment. *J Speech Hear Res* 1994;**37**(6):1414–23. Exclude: 7.

Christie J, McIntyre J, Burnham D. An evaluation of group parent training as an alternative method of treatment for articulation impairment. *Aust J Hum Commun Disord* 1991;**19**(1):3–22. Exclude: 7.

Click MJ, Ueberle JK, George CE. ESEA Title III project design for language-handicapped kindergarten and preschool children. *Lang Speech Hear Serv Schools* 1973;**4**:174–81. Exclude: 7.

Cole KN, Dale PS. Direct language instruction and interactive language instruction with language delayed preschool children: a comparison study. *J Speech Hear Res* 1986;**29**(2):206–17. Exclude: 7.

Connell PJ. Acquisition of semantic role by language-disordered children: differences between production and comprehension. *J Speech Hear Res* 1986;**29**:366–74. Exclude: 7.

Connell PJ. An effect of modeling and imitation teaching procedures on children with and without specific language impairment. *J Speech Hear Res* 1987;**30**(1):105–13. Exclude: 7.

Cook J, Urwin S, Kelly K. Preschool language intervention – a follow-up of some within-group differences. *Child Care Health Dev* 1989;**15**(6):381–400. Exclude: 7.

Cooper J, Moodley M, Reynell J. The developmental language programme: results from a five year study. *Br J Disord Commun* 1979;**14**(1):57–69. Exclude: 6.

Costello J, Onstine JM. The modification of multiple articulation errors based on distinctive feature theory. *J Speech Hear Disord* 1976;**41**(2):199–215. Exclude: 7.

Costello J, Schoen J. The effectiveness of para-professionals and speech clinicians as agents of articulation intervention using programmed instruction. *Lang Speech Hear Serv Schools* 1978;**9**:118–28. Exclude: 7.

Courtright JA, Courtright IC. Imitative modeling as a language intervention strategy: the effects of two mediating variables. *J Speech Hear Res* 1979;**22**:389–402. Exclude: 7.

Cox J, Hill S. Tackling language delay: a groupwork approach. *Health Visitor* 1993;**66**(8):291–2. Exclude: 6 7.

Dodd B, Barker R. The efficacy of utilizing parents and teachers as agents of therapy for children with phonological disorders. *Aust J Hum Commun Disord* 1990;**18**(1):29–45. Exclude: 7.

Dodd B, Iacano T. Phonological disorders in children: changes in phonological process use during treatment. *Br J Disord Commun* 1989;**24**:333–51. Exclude: 7.

Dukes PJ. An exploratory study of the comparative effectiveness of two language intervention programs in teaching normal and language deviant preschool children [PhD thesis]. Kent State University, 1974. Exclude: 7.

Dunn C, Barron C. A treatment program for disordered phonology: phonetic and linguistic considerations. *Lang Speech Hear Serv Schools* 1982;**13**(2):100–9. Exclude: 7.

Eiserman WD, Weber C, McCoun M. Two alternative program models for serving speech-disordered preschoolers: a second year follow-up. *J Commun Disord* 1992;**25**(2–3):77–106. Exclude: 7.

Elbert M, McReynolds LV. Aspects of phonological acquisition during articulation training. *J Speech Hear Disord* 1979;**44**(4):459–71. Exclude: 7.

Elbert M, McReynolds LV. The generalization hypothesis: final consonant deletion. *Lang Speech* 1985;**28**(3):281–94. Exclude: 7.

Evans CM, Potter RE. The effectiveness of the 's' pack when administered by 6th grade children to primary grade children. *Lang Speech Hear Serv Schools* 1974;**5**:85–90. Exclude: 7.

Evesham M. Teaching language skills to children with language disorders. *Br J Disord Commun* 1977;**12**(1):23–9. Exclude: 7.

Fey ME, Stalker CH. A hypothesis-testing approach to treatment of a child with an idiosyncratic (morpho) phonological system. *J Speech Hear Disord* 1986;**51**(4):324–36. Exclude: 7.

Fey ME, Cleave PL, Long SH. Two models of grammar facilitation in children with language impairments: Phase 2. *J Speech Hear Res* 1997;**40**:5–19. Exclude: 7.

Forrest K, Weismer G, Hodge M, Dinnseny DA, Elbert M. Statistical analysis of word – initial /k/ and /t/ produced by normal and phonologically disordered children. *Clin Linguist Phonetics* 1990;**4**:327–40. Exclude: 7.

Frattali CM. Quality of care in speech-language pathology: treatment evaluation of articulation disordered children in a university speech and hearing clinic. 1987 (Unpublished). Exclude: 7.

Friedman P, Mulhern ST. Relationship of clinician feedback to child initiated verbalization during language training. *J Commun Disord* 1976;**9**(4):289–99. Exclude: 7.

Gibbon F, Shockey L, Reid J. Description and treatment of abnormal vowels in a phonologically disordered child. *Child Lang Teaching Ther* 1992;**8**(1):30–59. Exclude: 7.

Haden R, Penn C. The twin situation and its effects on syntax and interactional language over time. *Br J Disord Commun* 1985;**20**(1):19–29. Exclude: 7.

Haley KL, Camarata SM, Nelson KE. Social valence in children with specific language impairment during imitation-based and conversation-based language intervention. *J Speech Hear Res* 1994;**37**(2):378–88. Exclude: 7.

Hoffman PR, Norris JA, Monjure J. Comparison of process targeting and whole language treatments for phonologically delayed preschool children. *Lang Speech Hear Serv Schools* 1990;**21**:102–9. Exclude: 7.

Hooshyar N. Language interactions between mothers and their nonhandicapped children, mothers and their Down Syndrome children, mothers and their language-impaired children. *Int J Rehab Res* 1985;**8**(4):475–77. Exclude: 7.

Howell J, Dean E. Does metaphon therapy work? In: Howell J, Dean E, editors. Treating phonological disorders in children: metaphon – theory to practice. 2nd ed. London: Whurr Publishers Ltd, 1991:151–70. Exclude: 7.

Hurd A. The Netherton concept project: a collaborative approach to assessment and teaching. *Child Lang Teaching Ther* 1990;**6**(1):33–40. Exclude: 7.

Kamhi AG, Gentry B, Mauer D, Gholson B. Analogical learning and transfer in language-impaired children. *J Speech Hear Disord* 1990;**55**(1):140–8. Exclude: 7.

Kemper RL. A parent-assisted early childhood environmental language intervention program. *Lang Speech Hear Serv Schools* 1980;**11**:229–35. Exclude: 7.

Koniditsiotis CY, Hunter TL. Speech-pathology-based language intervention within an early childhood education environment. *Early Child Dev Care* 1993;**96**:93–9. Exclude: 6 7.

Kot A, Law J. Intervention with preschool children with specific language impairments: a comparison of two different approaches to treatment. *Child Lang Teaching Ther* 1995;**11**(2):144–62. Exclude: 7.

Kupperman P, Bligh S, Goodban M. Activating articulation skills through theraplay. *J Speech Hear Disord* 1980;**45**(4):540–8. Exclude: 7.

Leahy J, Dodd B. The development of disordered phonology: a case study. *Lang Cognitive Processes* 1987;**2**(2):115–32. Exclude: 7.

Le Normand MT, Chevrie-Muller C. A follow-up case study of transitory developmental apraxia of speech: 'L'enfant a voyelles'. *Clin Linguist Phonetics* 1991;**5**(2):99–118. Exclude: 7.

Leonard LB, Schwartz RG, Chapman K, Rowan LE, Prelock PA, Terrell B, *et al.* Early lexical acquisition in children with specific language impairment. *J Speech Hear Res* 1982;**25**:554–64. Exclude: 7.

Locke A. Screening and intervention with children with speech and language difficulties in mainstream schools. In: Mogford K, Sadler J, editors. Child language disability: implications in an educational setting. Clevedon, UK: Multilingual Matters, Ltd, 1989:40–51. Exclude: 7.

Lowenthal B. Effect of small-group instruction on language-delayed preschoolers. *Except Child* 1981;**48**(2):178–9. Exclude: 7.

Masterson JJ, Daniels DL. Motoric versus contrastive approaches to phonology therapy: a case study. *Child Lang Teach Ther* 1991;**7**(2):127–40. Exclude: 7.

McReynolds LV, Elbert MF. Generalization of correct articulation in clusters. *Appl Psycholinguist* 1981;**2**(2):119–32. Exclude: 7.

Mulac A, Tomlinson CN. Generalization of an operant remediation program for syntax with language delayed children. *J Commun Disord* 1977;**10**(3):231–43. Exclude: 7.

Navratil K, Petrasek M. An approach to mainstreaming language-delayed children in the elementary school. *Lang Speech Hear Serv Schools* 1978;**9**:8–16. Exclude: 5 6 7.

Nelson KE, Camarata SM, Welsh J, Butkovsky L, Camarata M. Effects of imitative and conversational recasting treatment on the acquisition of grammer in children with specific language impairment and younger language-normal children. *J Speech Hear Res* 1996;**39**:850–9. Exclude: 7.

Oetting JB, Rice ML, Swank LK. Quick incidental learning (QUIL) of words by school-age children with and without SLI. *J Speech Hear Res* 1995;**38**(2):434–45. Exclude: 7.

Onslow DJ. The efficacy of intensive group speech therapy for inner city areas. *Child Lang Teaching Ther* 1988;**4**(1):26–45. Exclude: 5 6 7.

Paul-Brown D. A classroom-based model of language intervention for preschool language-impaired children: principles and procedures. *Ann Dyslexia* 1988;**38**:193–207. Exclude: 7.

Ruder KF, Bunce BH. Articulation therapy using distinctive feature analysis to structure the training program: two case studies. *J Speech Hear Disord* 1981;**46**(1):59–65. Exclude: 7.

Ruscello DM, Cartwright LR, Haines KB, Shuster LI. The use of different service delivery models for children with phonological disorders. *J Commun Disord* 1993;**26**(3):193–203. Exclude: 7.

Rvachew S. Speech perception training can facilitate sound production learning. *J Speech Hear Res* 1994;**37**(2):347–57. Exclude: 7.

Saben CB, Ingham JC. The effects of minimal pairs treatment on the speech-sound production of two children with phonologic disorders. *J Speech Hear Res* 1991;**34**(5):1023–40. Exclude: 7.

Shelton RL, Willis V, Johnson AF, Arndt WB. Oral form recognition training and articulation change. *Percept Mot Skills* 1973;**36**(2):523–31. Exclude: 7.

Shriberg LD, *et al.* Tabletop versus microcomputer-assisted speech management: response evocation phase. *J Speech Hear Disord* 1990;**55**(4):635–55. Exclude: 7.

Shriberg LD, Kwiatkowski J. Self-monitoring and generalization in preschool speech-delayed children. *Lang Speech Hear Serv Schools* 1990;**21**(3):157–70. Exclude: 7.

Stackhouse J, Snowling M. Developmental verbal dyspraxia II: a developmental perspective on two case studies. *Eur J Disord Commun* 1992;**27**(1):35–54. Exclude: 7.

Stone CA, Connell PJ. Induction of a visual symbolic rule in children with specific language impairment. *J Speech Hear Res* 1993;**36**(3):599–608. Exclude: 7.

Swisher L, Restrepo MA, Plante E, Lowell S. Effect of implicit and explicit 'rule' presentation on bound-morpheme generalization in specific language impairment. *J Speech Hear Res* 1995;**38**(1):168–73. Exclude: 7.

Tyler AA, Edwards ML, Saxman JH. Clinical application of two phonologically based treatment procedures. *J Speech Hear Disord* 1987;**52**(4):393–409. Exclude: 7.

Tyler AA, Edwards ML, Saxman JH. Acoustic validation of phonological knowledge and its relationship to treatment. *J Speech Hear Disord* 1990;**55**(2):251–61. Exclude: 7.

Tyler AA, Watterson KH. Effects of phonological versus language intervention in preschoolers with both phonological and language impairment. *Child Lang Teaching Ther* 1991;**7**:141–60. Exclude: 7.

Velleman SL. The interaction of phonetics and phonology in developmental verbal dyspraxia: two case studies. *Clin Commun Disord* 1994;**4**(1):66–77. Exclude: 7.

Weiner F. Treatment of phonological disability using the method of meaningful minimal contrast: two case studies. *J Speech Hear Disord* 1981;**46**:91–6. Exclude: 7.

Wilcox MJ, Kouri T, Caswell S. Early language intervention. *Am J Speech Lang Pathol* 1991;49–61. Exclude: 7.

Williams GC, McReynolds LV. The relationship between discrimination and articulation training in children with misarticulations. *J Speech Hear Res* 1975;**18**(3):401–12. Exclude: 7.

Wing CS. A preliminary investigation of generalization to untrained words following two treatments of children's word-finding problems. *Lang Speech Hear Serv Schools* 1990;**21**:151–6. Exclude: 7.

D. Screening

Appelbaum AS. Validity of the revised Denver Developmental Screening Test for referred and nonreferred samples. *Psychol Rep* 1978;**43**(1):227–33. Exclude: 8 9.

Bax M, Whitmore K. Neurodevelopmental screening in the school-entrant medical examination. *Lancet* 1973;**2**(825):368–70. Exclude: 8 9.

Beilby P. Newham Speech Therapy Department's Screening Device. 1987 (Unpublished). Exclude: 8 9 10.

Bondestam M, Hagelin E, Rydell A-M, Westerlund M. Health examination of preschool children in relation to school adjustment in grade 1. *Acta Paediatr* 1992;**81**:257–61. Exclude: 8.

Bowers R, Oakenfull S. The role of health visitors in speech and language therapy. *Health Visitor* 1996;**69**(8):319–20. Exclude: 6–10.

Bricker D, Squires J. Low cost system using parents to monitor the development of at-risk infants. *J Early Intervention* 1989;**13**(1):50–60. Exclude: 8 9 10.

Bricker D, Squires J. The effectiveness of parental screening of at-risk infants: the infant monitoring questionnaires. *Top Early Child Spec Ed* 1989;**9**:67–85. Exclude: 8 9.

Bryant GM, Davies KJ, Newcombe RG. The Denver Developmental Screening Test: achievement of test items in the first year of life by Denver and Cardiff infants. *Dev Med Child Neurol* 1974;**16**:475–84. Exclude: 9 10.

Burgess DB, Asher KN, Doucet HJ, Reardon K, Daste MR. Parent report as a means of administering the Prescreening Developmental Questionnaire: an evaluation study. *J Dev Behav Pediatr* 1984;**5**(4):195–200. Exclude: 8 9.

Byrne J, Ellsworth C, Bowering E, Vincer M. Language development in low birth weight infants: the first two years of life. *J Dev Behav Pediatr* 1993;**14**(1):21–7. Exclude: 7.

Byrne JM, Backman JE, Bawden HN. Minnesota Child Development Inventory: a normative study. *Can Psychol* 1995;**36**(2):115–30. Exclude: 8 9 10.

Byrne JM, Backman JE, Smith IM. Developmental assessment: the clinical use and validity of parental report. *J Pediatr Psychol* 1986;**11**(4):549–59. Exclude: 8 9.

Cadman D, Chambers LW, Walter SD, Feldman W, Smith K, Ferguson R. The usefulness of the Denver Developmental Screening Test to predict kindergarten problems in a general community population. *Am J Public Health* 1984;**74**(10):1093–7. Exclude: 8 9.

Cadman D, Walter SD, Chambers LW, Ferguson R, Szatmari P, Johnson N, *et al.* Predicting problems in school performance from preschool health, developmental and behavioural assessments. *Can Med Assoc J* 1988;**139**(1):31–6. Exclude: 8 9.

Camaioni L, Caselli MC, Longobardi E, Volterra V. Costruzione e validazione di uno strumento per rilevare lo sviluppo comunicativo-linguistico nel secondo anno di vita. [Construction and validation of an instrument for evaluating communicative-linguistic development during the second year of life.]. *Giornale Italiano di Psicologia* 1991;**18**(3):419–37. Exclude: 8 9 10.

Camaioni L, Castelli MC, Longobardi E, Volterra V. A parent report instrument for early language assessment. *First Lang* 1991;**11**(33 Pt 3):345–59. Exclude: 8 9 10.

Camp BW, van Doorninck WJ, Frankenburg WK, Lampe JM. Preschool developmental testing in prediction of school problems. *Clin Pediatr* 1977;**16**(3):257–63. Exclude: 8 9.

Capute AJ, Accardo PJ. Linguistic and auditory milestones during the first two years of life. *Clin Pediatr* 1978;**17**(11):847–53. Exclude: 6–10.Capute AJ, Palmer FB, Shapiro BK, Wachtel RC, Schmidt S, Ross A. Clinical Linguistic and Auditory Milestone Scale: prediction of cognition in infancy. *Dev Med Child Neurol* 1986;**28**(6):762–71. Exclude: 8 9 10.

Capute AJ, Shapiro BK, Palmer FB. Marking the milestones of language development. *Contemp Pediatr* 1987;**4**:24–41. Exclude: 6–10.

Capute AJ, Shapiro BK, Wachtel RC, Gunther VA, Palmer FB. The Clinical Linguistic and Auditory Milestone Scale (CLAMS). Identification of cognitive deficits in motor-delayed children. *Am J Dis Child* 1986;**140**:694–8. Exclude: 8 9.

Cartwright LR, Lass NJ. A comparative study of children's performance on the Token Test, Northwestern Syntax Screening Test, and Peabody Picture Vocabulary Test. *Acta Symbolica* 1974;**5**(1):19–29. Exclude: 8 9 10.

Cole KN, Fewell RR. A quick language screening test for young children: The Token Test. *J Psychoed Assess* 1983;**1**(2):149–53. Exclude: 10.

Colligan RC. Prediction of reading difficulty from parental report: a three year follow-up. *Learn Disab Quart* 1981;**4**(1):31–7. Exclude: 8 9.

Colligan RC. Prediction of kindergarten reading success from preschool report of parents. *Psychol Sch* 1976;**13**(3):304–8. Exclude: 8 9 10.

Cools AT, Hermanns JM. Vroegtijdige onderkenning van problemen in de ontwikkeling van kinderen: de Denver Ontwikkeling Screeningtest (DOS). [Early identification in the development of children: the Denver Developmental Screening Test (DDST)]. *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden* 1976;**31**(4):179–200. Exclude: 8 9 10.

Coplan J, Gleason JR. Quantifying language development from birth to three years using the Early Language Milestone Scale. *Pediatrics* 1990;**86**(6):963–71. Exclude: 8 9 10.

Coplan J, Gleason JR. Unclear speech: recognition and significance of unintelligible speech in preschool children [published erratum appears in *Pediatrics* 1988;**82**(6):862]. *Pediatrics* 1988;**82**(3 Pt 2):447–52. Exclude: 8.

Creighton DE, Sauve RS. The Minnesota Infant Development Inventory in the developmental screening of high-risk infants at eight months. Special Issue: Child and adolescent health. *Can J Behav Sci* 1988;**20**(4):424–33. Exclude: 7.

Dale PS. The validity of a parent report measure of vocabulary and syntax at 24 months. *J Speech Hear Res* 1991;**34**(3):565–71. Exclude: 8 10.

Dale PS, Bates E, Reznick JS, Morisset C. The validity of a parent report instrument of child language at twenty months. *J Child Lang* 1989;**16**(2):239–49. Exclude: 8 10.

Damico J, Oller JW, Jr. Pragmatic versus morphological/syntactic criteria for language referrals. *Lang Speech Hear Serv Sch* 1980;**11**(2):85–94. Exclude: 10.

Dearlove J, Kearney D. How good is general practice developmental screening? *BMJ* 1990;**300**:1177–80. Exclude: 8 9.

Diamond KE. Predicting school problems from preschool developmental screening: a four-year follow-up of the Revised Denver Developmental Screening Test and the role of parent report. *J Div Early Child* 1987;**11**(3):247–53. Exclude: 8 9.

Drillien CM, Pickering RM, Drummond MB. Predictive value of screening for different areas of development. *Dev Med Child Neurol* 1988;**30**:294–305. Exclude: 9 10.

Duncan JC, Perozzi JA. Concurrent validity of a pragmatic protocol. *Lang Speech Hear Serv Sch* 1987;**18**(1):80–5. Exclude: 10.

Egan DF, Brown R. Developmental assessment: 18 months to $4^{1}/_{2}$ years. The Bus Puzzle Test. *Child Care Health Dev* 1984;**10**(3):163–79. Exclude: 8 9 10.

Eno L, Woehlke P. Predicting preschool speech/language referral status with the Lollipop Test and the Cognitive-Language Profile of the Early Screening Profiles. *Percept Mot Skills* 1995;**80**(1):1025–6. Exclude: 10.

Epir S, Yalaz K. Urban Turkish children's performance on the Denver Developmental Screening Test. *Dev Med Child Neurol* 1984;**26**(5):632–43. Exclude: 8 9 10.

Eu BSL. Evaluation of a developmental screening system for use by child health nurses. *Arch Dis Child* 1986;**61**:34–41. Exclude: 10.

Fenson L, Dale PS, Reznick JS, Bates E, Thal DJ, Pethick SJ, *et al.* Variability in early communicative development. *Monogr Soc Res Child Dev* 1994;**595**(242):1–185. Exclude: 10.

Fisher J. ZAP2: a screening test for young school children. *Child Lang Teaching Ther* 1993;**9**(2):142–51. Exclude: 8 9 10.

Fluharty NB. The design and standardisation of a Speech and Language Screening Test for use with preschool children. *J Speech Hear Disord* 1974;**39**(1):75–87. Exclude: 10.

Frankenburg WK, Camp BW, *et al.* The reliability and stability of the Denver Developmental Screening Test. *Child Dev* 1971;**42**:1315–25. Exclude: 8 9 10.

Frankenburg WK, Camp BW, van Natta PA. Validity of the Denver Developmental Screening Test. *Child Dev* 1971;**42**:475–88. Exclude: 8 9.

Frankenburg WK, Dodds J, Archer P, Shapiro H, Bresnick B. The Denver II: a major revision and restandardization of the Denver Developmental Screening Test. *Pediatrics* 1992;**89**(1):91–7. Exclude: 8 9 10.

Frankenburg WK, Dodds JB. The Denver Developmental Screening Test. *Pediatrics* 1967;**71**(2):181–91. Exclude: 8 9.

Frankenburg WK, *et al.* The Denver Pre-screening Developmental Questionnaire. *Pediatrics* 1976;**57**(5):744–53. Exclude: 8 9.

Frankenburg WK, Goldstein AD, Camp BW. The Revised Denver Developmental Screening Test: its accuracy as a screening instrument. *J Pediatr* 1971;**79**(6):988–95. Exclude: 8 9.

Gale M, Holloway K and the Southmead and Frenchay Speech and Language Therapy Services. BRISC: The Bristol Surveillance of Children's Communication. 2nd ed. 1995. Exclude: 6–10.

Gibbon F, Ford E. Pilot results indicate need for more research. *Speech Ther Practice* 1989;July 3–4. Exclude: 8 9 10.

Gibbon F, Ford E. An agenda for screening three year olds. *Speech Ther Practice* 1989;June 4–5. Exclude: 8 9 10.

Gillham B, Boyle J, Smith N, Cheyne B. Language Screening at 18–36 months: the First Words and First Sentences Tests and the First Words Comprehension Cards. *Child Lang Teaching Ther* 1995;**11**(2):193–207. Exclude: 6–10.

Glascoe FP. Parents' concerns about children's development: prescreening technique or screening test? *Pediatrics* 1997;**99**(4):522–8. Exclude: 10.

Glascoe FP. It's not what it seems. The relationship between parents' concerns and children with global delays. *Clin Pediatr* 1994;**33**:292–6. Exclude: 8 9.

Glascoe FP, Altemeier WA, MacLean WE. Are parents' concerns valid indicators of their child's development? *Am J Dis Child* 1988;**142**:383. Exclude: 8 9.

Glascoe FP, Altemeier WA, MacLean WE. The importance of parents' concerns about their child's development. *Am J Dis Child* 1989;**143**(8):955–8. Exclude: 9.

Glascoe FP, Byrne K, Ashford L, Johnson K, Chang B, Strickland B. Accuracy of the Denver II in developmental screening. *Pediatrics* 1992;**89**(2):1221–5 [Comments in *Pediatrics* 1992;**90**(6):1009–11]. Exclude: 9.

Glascoe FP, Byrne KE. The usefulness of the Battelle Developmental Inventory Screening Test. *Clin Pediatr* 1993;**32**(5):273–80. Exclude: 8 9.

Glascoe FP, Byrne KE. The usefulness of the Developmental Profile – II in developmental screening. *Clin Pediatr* 1993;**32**(4):203–8 [Comments in *Clin Pediatr* 1993;**32**(4):209–12]. Exclude: 9.

Goldberg R. Identifying speech and language delays in children. *Pediatr Nurs* 1984;**10**:252–9. Exclude: 6–10.

Goorhuis-Brouwer SM. Fruhzeitige Erkennung von Sprachentwicklungsstorungen: Die Eltern als engagierte Beobachter von Sprachentwicklungsstorungen [Early detection of language disorders: the parents as observers of language development]. *Folia Phoniatrica* 1990;**42**(5):260–4. Exclude: 9 10.

Gorrell RW. The validity and predictive efficiency of the North Carolina Psychoeducational Screening Test. *J Pediatr Psychol* 1981;**6**(4):435–39. Exclude: 8 9.

Gottfried AW, Guerin D, Spencer JE, Meyer C. Concurrent validity of the Minnesota Child Development Inventory in a nonclinical sample. *J Consult Clin Psychol* 1983;**51**(4):643–4. Exclude: 8 9.

Gottfried AW, Guerin D, Spencer JE, Myer C. Validity of Minnesota Child Development Inventory in screening young children's developmental status. *J Pediatr Psychol* 1984;**9**(2):219–30. Exclude: 8 10.

Greenough P, Illerbrun D, Haines L, McLeod J. Is language-learning disability a life-span phenomenon? Kindergarten screening – a positive step. *Sch Psychol Int* 1984;**5**(4):218–23. Exclude: 6–10.

Gudmundsson E, Gretarsson SJ. Reliability and validity of mothers' developmental estimates for children between 4 and 41 months. *Scand J Psychol* 1994;**35**(4):336–42. Exclude: 8 9 10.

Gudmundsson EG, Gretarsson SJ. Reliability and validity of an Icelandic developmental parent report: children from 36–72 months. *Educ Psychol Measurement* 1993;**53**:571–83. Exclude: 10.

Haines CR, Brown JB, Grantham EB, Rajagopalan VS, Sutcliffe PV. Neurodevelopmental screen in the school entrant medical examination as a predictor of co-ordination and communication difficulties. *Arch Dis Child* 1985;**60**(12):1122–7. Exclude: 9 10.

Harris SR. Parents' and caregivers' perceptions of their children's development. *Dev Med Child Neurol* 1994;**36**:918–23. Exclude: 8 9.

Henderson LW, Meisels SJ. Parental involvement in the developmental screening of their young children: a multiple-source perspective. *J Early Intervention* 1994;**18**(2):141–54. Exclude: 8 9.

Henin N, Du Lac M. Examen de langage chez l'enfant de 18 mois a 6 ans et demi; Methodes, Reflexions. *Revue de Layyngologie* 1984;**105**(4):369–73. Exclude: 6–10.

Holmes GE, Hassanein RS. The KIDS Chart. A simple, reliable infant development screening tool. *Am J Dis Child* 1982;**136**(11):997–1001. Exclude: 8 9 10.

Hoon AH, Jr, Pulsifer MB, Gopalan R, Palmer FB, Capute AJ. Clinical Adaptive Test/Clinical Linguistic Ausitory Milestone Scale in early cognitive assessment. *J Pediatr* 1993;**123**:S1–S8. Exclude: 8 9.

Hurd A. The Netherton concept project: a collaborative approach to assessment and teaching. *Child Lang Teaching Ther* 1990;**6**(1):33–40. Exclude: 9 10.

Ireton H. The Child Development Review: Monitoring children's development using parents' and pediatricians' observations. *Infants Young Child* 1996;**9**(1):42–52. Exclude: 8 9 10.

Ireton H, Diamond G, Carney J. Preschool screening using the parents' preschool development inventory report. *Diagnostique* 1993;**18**(4):269–81. Exclude: 8 9.

Ireton H, Glascoe FP. Assessing children's development using parents' reports. *Clin Pediatr* 1995;**34**(5):248–55. Exclude: 8 9 10.

Ireton H, Thwing E, Currier SK. Minnesota Child Development Inventory: identification of children with developmental disorders. *J Pediatr Psychol* 1977;**2**(1):18–22. Exclude: 9.

Jaffe M, Havel J, Goldberg A, Rudolph-Schnitzer M, Winter ST. The use of the Denver Developmental Screening Test in infant welfare clinics. *Dev Med Child Neurol* 1980;**22**(1):55–60. Exclude: 8 9.

James DGH. The efficacy of a language screening battery for Australian five-year old children. *Aust J Hum Commun Disord* 1994;**22**(2):31–46. Exclude: 10.

Johnston O. Ill health and developmental delays in Adelaide four-year-old children. *Aust Paediatr J* 1980;**16**(4):248–54. Exclude: 8 9 10.

Keenan PA, Lachar D. Screening preschoolers with special problems: use of the Personality Inventory for Children (PIC). *J Sch Psychol* 1988;**26**(1):1–11. Exclude: 9.

Kelly DJ, Rice ML. A strategy for language assessment of young children: a combination of two approaches. *Lang Speech Hear Serv Sch* 1986;**17**:83–94. Exclude: 6–10.

Kenny TJ, Hebel JR, Sexton MJ, Fox NL. Developmental screening using parent report. *J Dev Behav Pediatr* 1987;**8**(1):8–11. Exclude: 8 9.

Kitson DL, Vance B, Blosser J. Comparison of the Token Test of Language Development and the Wechsler Intelligence Scale for Children – Revised. *Percept Mot Skills* 1985;**61**(2):532–4. Exclude: 8 9 10.

Knobloch H, Stevens F, Malone A, Ellison P, Risemberg H. The validity of parent reporting of infant development. *Pediatrics* 1979;**63**(6):872–8. Exclude: 8 9.

Kopparthi R, McDermott C, Sheftel D, Connors-Lenke M, Getz M, Frey M. The Minnesota Child Development Inventory: validity and reliability for assessing development in infancy. *J Dev Behav Pediatr* 1991;**12**(4):217–22 [Comment in *J Dev Behav Pediatr* 1992;**13**(2):147–8]. Exclude: 8.

Kulig SG, Baker KA. Preliminary field-testing of The Physician's Developmental Quick Screen for Speech Disorders ('PDQ'). A 5-minute procedure for testing children six months to six years. *Clin Pediatr (Phila)* 1976;**15**(12):1146–50. Exclude: 10.

Labeck LJ, Ireton H, Leeper SD. On-line computerised assessment of young children using the Minnesota Child Development Inventory. *Child Psychiatry Hum Dev* 1983;**14**(1):49–54. Exclude: 6–10.

Lamorey SG. High-risk parents as developmental screeners: effects of parent-completed developmental questionnaires on parental knowledge of infant development and parent–infant interaction [Dissertation: Division of Special Education and Rehabilitation and the graduate school of the University of Oregon, in partial fulfillment of the requirements for the degree of doctor of philosophy], 1992. Exclude: 8 9 10.

Larson GW, Summers PA. Response patterns of preschool-age children to the Northwestern Syntax Screening Test. *J Speech Hear Disord* 1976;**41**(4):486–97. Exclude: 8 9 10.

Lichtenstein R. Predicting school performance of preschool children from parent reports. *J Abnorm Child Psychol* 1984;**12**(1):79–93. Exclude: 8 9 10.

Lipson AM. Catching them early. *Acad Ther* 1981;**16**(4):457–62. Exclude: 6–10.

Llewellyn G, Maher L. Assessment of school students' needs for therapy services. *Aust J Physiother* 1993;**39**(3):181–5. Exclude: 6–10.

Locke A. Screening and intervention with children with speech and language difficulties in mainstream schools. In: Mogford K, Sadler J, editors. Child language disability: implications in an educational setting. Clevedon, UK: Multilingual Matters, Ltd, 1989:40–51. Exclude: 6–10.

Lynch J. Use of a prescreening checklist to supplement speech-language-hearing screening. *Lang Speech Hear Serv Sch* 1979;**10**(4):249–58. Exclude: 6–10.

Magill S, Bell M, McKee A, Devon C. A collaborative model of speech and language therapy for children. Presentation for Hewlett-Packard Golden Helix Award. 1996, (Unpublished). Exclude: 6–10.

Mathur GP, Mathur S, Singh YD, Kushwaha KP, Lele SN. Detection and prevention of childhood disability with the help of Anganwadi workers. *Indian Pediatr* 1995;**32**(7):773–7. Exclude: 8 9 10.

McCartney K, Scarr S, Goldfarb A, Hrncir E. The screening, assessment and treatment of language delay in young children: an evaluation plan. *First Lang* 1983;**4**(10 Pt 1):63–5. Exclude: 6–10.

Miller LJ, Schouten PG. Age-related effects on the predictive validity of the Miller Assessment for Preschoolers. *J Psychoeduc Assess* 1988;**6**(2):99–106. Exclude: 8 9 10.

Moe JC. The reliability and validity of the Minnesota Child Development Inventory for assessing the developmental status of moderately retarded children [PhD thesis]. Kent State University, 1984. Exclude: 8 10.

Monsees EK, Berman C. Speech and language screening in a summer headstart program. *J Speech Hear Disord* 1968;**33**:121–6. Exclude: 6–10.

Morgan RL, Dawson B, Kerby D. The performance of preschoolers with speech/language disorders on the McCarthy Scales of Children's Abilities. *Psychol Sch* 1992;**29**(1):11–17. Exclude: 9 10.

Mott SE. Concurrent validity of the Battelle Developmental Inventory for speech and language disordered children. *Psychol Sch* 1987;**24**(3):215–20. Exclude: 8 9 10.

O'Byran KG. The Windsor Early Identification Project revisited. Early identification and intervention: 4th Symposium on Learning Problems. 1979 Toronto. Exclude: 8 9 10.

Olade RA. Evaluation of the Denver Developmental Screening Test as applied to African children. *Nurs Res* 1984;**33**(4):204–7. Exclude: 8 9 10.

Pulsifer MB, Hoon AH, Palmer FB, Gopalan R, Capute AJ. Maternal estimates of developmental age in preschool children. *J Pediatr* 1994;**125**(1):S18–S24. Exclude: 8 9.

Radway CJ, Norfolk MB. The Surrey Speech Screening Test. *Medical Officer* 1971;April:185–7. Exclude: 8.

Ratusnik DL, Klee TL, Melnick Ratusnik C. Northwestern Syntax Screening Test: a short form. *J Speech Hear Disord* 1980;**45**(2):200–8. Exclude: 8 9 10.

Ratusnik DL, Koenigsknecht RA. Internal consistency of the Northwestern Syntax Screening Test. *J Speech Hear Disord* 1975;**40**(1):59–68. Exclude: 8 9 10.

Ratusnik DL, Ratusnik CM, Sattinger K. Short-form application of the Screening Test of Spanish Grammar and the Northwestern Syntax Screening Test to Latino children. *Lang Speech Hear Serv Sch* 1978;**9**(4):220–35. Exclude: 8 9 10.

Renfrew CE, Geary L. Prediction of persisting speech deficit. *Br J Disord Commun* 1973;**8**(1):37–41. Exclude: 8 9 10.

Reznick JS, Goldsmith L. A multiple form word production checklist for assessing early language. *J Child Lang* 1989;**16**(1):91–100. Exclude: 8 9 10.

Riley GD, Riley J. Physician's screening procedure for children who may stutter. *J Fluency Disord* 1989;**14**:57–66. Exclude: 8 9 10.

Ritter DR. Preschool attainment record as a measure of developmental skills. *J Consult Clin Psychol* 1977;**45**(6):1184. Exclude: 8 9 10.

Ritterman SI, Zook-Herman S, Carlson RL, Kinde SW. The pass/fail disparity among three commonly employed articulation screening tests. *J Speech Hear Disord* 1982;**47**(4):429–33. Exclude: 8 9 10.

Roman VP. The relationship between language ages of preschool children derived from a parent informant scale and language ages derived from tests administered directly to the preschool child. *Lang Speech Hear Serv Sch* 1980;**11**(1):50–5. Exclude: 10.

Rona RJ, Allsop M, Morris RW. Monitoring the developmental examination schedule in an inner city district. *Child Care Health Dev* 1987;**13**:329–40. Exclude: 8 9 10.

Rossman MJ, Hyman SL, Rorabaugh ML, Berlin LE, Allen MC, Modlin JF. The CAT/CLAMS assessment for early intervention services. *Clin Pediatr* 1994;**33**(7):404–9. Exclude: 8 9.

Roy B, Maeder C. Validation de ERTL4: Epreuve de reperage des troubles du langage lors du bilan medical de quatre ans [ERTL4: Screening test for detecting language disorders during medical examination of four year old children]. Conference of the Comite Permanente de Liaison des Othophonistes-Logopedes del'UE [CPLOL], Lisbon, 1997. Exclude: 8.

Ruben RJ. Language screening as a factor in the management of the pediatric otolaryngic patient. *Arch Otolaryngol Head Neck Surg* 1991;**117**:1021–5. Exclude: 8 9.

Rydell A-M, Bondestam M, Hagelin E, Westerlund M. Teacher rated problems and school ability tests in relation to preschool problems and parents' health information at school start: a study of first graders. *Scand J Psychol* 1991;**32**:177–90. Exclude: 8 9.

Sanger DD, Aspedon M, Hux K, Chapman A. Early referral of school-age children with language problems. *J Child Commun Disord* 1995;**16**(2):3–9. Exclude: 9 10.

Saylor CF, Brandt BJ. The Minnesota Child Development Inventory: a valid maternal-report form for assessing development in infancy. *J Dev Behav Pediatr* 1986;**7**(5):308–11. Exclude: 8 9 10.

Schetz KF. Comparison of the Compton Speech and Language Screening Evaluation and the Fluharty Preschool Speech and Language Screening Test. *Lang Speech Hear Serv Sch* 1985;**16**(1):16–24. Exclude: 8 9 10.

Schwartz A, Murphy M. Cues for screening language disorders in preschool children. *Pediatrics* 1975;**55**:717–22. Exclude: 6–10.

Senecal J, Roussey M, Letort F, Delahaye M, Subileau B. Les causes d'inadaptation scolaire chez l'enfant. Depistage en maternelle et evolution. [Screening for school difficulties in kindergarten children]. *Archives Francaises de Pediatrie* 1987;**44**(7):523–30. Exclude: 8 9 10.

Shoemaker OS, Saylor CF, Erickson MT. Concurrent validity of the Minnesota Child Development Inventory with high risk infants. *J Pediatr Psychol* 1993;**18**(3):377–88. Exclude: 7 8 9.

Showronek H, Marx H. Die Bielefelder Langsschnittstudie zur Fruherkennung von Risiken der Lese-Rechtschreibschwache: Theoretischer Hintergrund und erste Befunde. *Heilpadagogische Forschung* 1989;**15**(1):38–49. Exclude: 8 9 10.

Siegel LS, Cooper DC, Fitzhardinge PM, Ash AJ. The use of the Mental Development Index of the Bayley Scale to diagnose language delay in 2-year-old high-risk infants. *Infant Behav Dev* 1995;**18**(4):483–6. Exclude: 7.

Sonnander K. Parental developmental assessment of 18 month old children: reliability and predictive value. *Dev Med Child Neurol* 1987;**29**(3):351–62. Exclude: 8 9.

Steckol KF. Speech and language screening of children from birth to 6 years of age. *J Ky Med Assoc* 1988;**86**:503–6. Exclude: 6–10.

Sturner RA, *et al.* Adaptations of the Denver Developmental Screening Test. A study of preschool screening. *Pediatrics* 1982;**69**:346–50. Exclude: 8 9.

Sturner RA, Funk SG, Green JA. Simultaneous Technique for Acuity and Readiness Testing (START): further concurrent validation of an aid for developmental surveillance. *Pediatrics* 1994;**93**(1):82–8. Exclude: 8 9.

Sturner RA, Funk SG, Thomas PD, Green JA. An adaptation of the Minnesota Child Development Inventory for preschool developmental screening. *J Pediatr Psychol* 1982;**7**(3):295–306. Exclude: 8 9.

Sturner RA, Green JA, Funk SG. Preschool Denver Developmental Screening Test as a predictor of later school problems. *J Pediatr* 1985;**107**(4):615–21. Exclude: 8 9.

Sturner RA, Heller J, Funk S, Feezor M, Layton T, Lewis D. The Communication Screening System (CCS): preliminary validation of a potential replacement for independent screens for speech, language and hearing. 1996 (Unpublished). Exclude: 8.

Sundelin C, Vuille JC. Health screening of four year olds in a Swedish county: I. Organisation, methods, and participation. *Acta Paediatr Scand* 1975;**64**:795–800. Exclude: 8 9 10.

Sundelin C, Vuille JC. Health screening of four year olds in a Swedish county: II. Effectiveness in detecting health problems. *Acta Paediatr Scand* 1975;**64**:801–6. Exclude: 8 9 10.

Swift EW, Swift WJ, Camp BW, Silvern LW. Predictive value of early testing of auditory localization for language development. *Dev Med Child Neurol* 1981;**23**(3):306–12. Exclude: 9 10.

Telzrow CF. Longitudinal prediction of PPVT, VMI and reaction time performance by the Battelle Developmental Inventory and the Minnesota Child Development Inventory for three ages of preschool children [PhD thesis]. Kent State University, 1982. Exclude: 8 9 10.

Thomas J, Pozsonyi V and Epsom Health Care Trust. The Epsom Language Screening Assessment Project, 1996. Exclude: 8.

Thorburn MJ, Desai P, Durkin M. A comparison of efficacy of the key informant and community survey methods in the identification of childhood disability in Jamaica. *Ann Epidemiol* 1991;**1**:255–61. Exclude: 9.

Tomblin JB, Hardy JC, Hein HA. Predicting poor communication status in preschool children using risk factors present at birth. *J Speech Hear Res* 1991;**34**:1096–105. Exclude: 9 10.

Tomblin JB, Shonnock CM, Hardy JC. The concurrent validity of the Minnesota Child Development Inventory as a measure of young children's language development. *J Speech Hear Disord* 1989;**54**(1):101–5. Exclude: 8 9 10.

Ueda R. Standardization of the Denver Developmental Screening Test on Tokyo children. *Dev Med Child Neurol* 1978;**20**(5):647–56. Exclude: 8 9.

Ullman DG, Kausch DF. Early identification of developmental strengths and weaknesses in preschool children. *Except Child* 1979;**46**(1):8–13. Exclude: 8.

Wachtel RC, Shapiro BK, Palmer FB, Allen MC, Capute AJ. CAT/CLAMS. A tool for the pediatric evaluation of infants and young children with developmental delay. Clinical Adaptive Test/Clinical Linguistic and Auditory Milestone Scale. *Clin Pediatr* 1994;**33**(7):410–15. Exclude: 8 9 10.

Westerlund M. Fonologisk screening av 4-aringar [Phonological screening of 4 year olds: a clinical evaluation of speech, comprehension, grammar and need of help]. *Nordisk Tidsskrift for Logopedi og Foniatri* 1988;**1**:37–43. Exclude: 8.

Westerlund M. Inter-rater reliability of a screening instrument for identification of language impairment in young children. *Phoniatr Logopedic Progr Rep* 1996;**10**:3–8. Exclude: 8 9 10.

Westerlund M. Predictive power of a phonological screening test at four years of age in relation to later linguistic ability. *Scand J Logopedics Phoniatr* 1995;**20**:69–76. Exclude: 9.

Wong V, Lee PWH, Mak-Lieh F, Yeung CY, Leung PWL, Luk SL, *et al.* Language screening in preschool Chinese children. *Eur J Disord Commun* 1992;**27**(3):247–64. Exclude: 10.

Zaman SS, Khan NZ, Islam S, Banu S, Dixit S, Shrout P, *et al.* Validity of the 'Ten questions' for screening serious childhood disability: results from urban Bangladesh. *Int J Epidemiol* 1990;**19**(3):613–20. Exclude: 9.

Ziermann U. Differentielle Entwicklungsdiagnostik im 3. Lebensjahr: Verfahrensdarstellung [Differential developmental diagnostics in the 3rd year: procedure presentation]. *Psychologie fur die Praxis* 1988;**6**:61–9. Exclude: 8 9 10.

Zimmerman DR, Rhoads FA, Rothbaum PA, Feierstein C, Russo E, Rescorla L. Screening for language delay in two-year-olds in an inner city clinic. *Ambulatory Child Health* 1997;**2**:339–47. Exclude: 8.

Studies not located in time

Albritton TP, editor. Secondary speech-language problems: strategies for a new delivery system. American Speech-Language-Hearing Association Annual Convention, 1984; San Francisco, CA.

Aylward GP, Verhulst SJ, Bell S. The 18-month Early Neuropsychologic Optimality Rating Scale (ENORS-18): a predictive assessment instrument. *Dev Neuropsychol* 1988;**4**(1):47–61.

Bohme G, Gollnitz G. [Rehabilitation ability of patients with speech disorders.] *Deutsche Gesundheitsw* 1968;**23**(31):1476–8.

Bruce B, Sahlen B. Can severe language disorders be identified before age 3? *Logoped Phoniatr Vocol* 1996;**21**(1):37–42.

Buekers R, Hendriks JJT. Vroegtijdige diagnostiek en hulpverlening aan kinderen met een taalontwikkelingsstoornis. *Ned Tijdschr Geneeskd* 1996;**140**(35).

Clark JD, Schalik MC. Predicting articulatory proficiency of Second Grade Children. *Ohio J Speech Hear* 1976;**11**(2):90–100.

Debray M, Mattlinger MJ, Dugas M, Mousset MR. [Diagnosis and treatment of language retardation.] *Cahiers Medicaux* 1986;**11**:85–9.

De Ridder JG. Vroegtijdige onderkenning van Communciatieve ontwikkelingsstoornissen: ontwikkeling en evaluatie van het VTO-taalsignaleringsinstrument 0–3 jarigen, 1990 Rijksuniversiteit Leiden, Meppel Repro.

Evans JS, Bangs T. Effects of preschool language training on later academic achievement of children with language and learning disabilities: a descriptive analysis. *J Learn Disab* 1972;**5**(10):585–92.

Fergusson DM, Horwood LJ, Lynskey MT. A longitudinal study of early childhood education and subsequent academic achievement. *Aust Psychol* 1994;**29**(2):110–15.

Franc S, Gerard CL. Suivi longitudinal d'une population de sujects dysphasiques. *Approche Neuropsychologique des Apprentissages chez l'Enfant* 1996;**8**(37):36–40.

Genovese E, Trevisi P. Screening e diagnosi precoce delle patologie del linguaggio nella prima infanzia: test di valutazione. [Screening and early diagnosis of language disorders in infancy: test evaluation]. *Acta Phoniatr Latina* 1990;**12**(2):351–8.

Genovese E, Trevisi P. Scale di valutazione del linguaggio nel bambino da zero a tre anni. [Language evaluation scales for newborn to age three children]. *Acta Phoniatr Latina* 1990;**12**(4):515–20.

Glascoe FP, Borowitz KC. Improving the sensitivity of the language sector of the Denver Developmental Screening Test. *Diagnostique* 1988;**13**(2–4):76–85.

Herbert GW, Wedell K. Communication handicaps of children with specific language deficiency. Annual Conference of the British Psychological Society, 1970; Southampton, UK.

Huskey R, Knight N, Oltman S, Irwin JV. A longitudinal study of the spontaneous remission of articulatory defects of 1665 school children in grades 1, 2 and 3. Part I: Reliability. *Acta Symbolica* 1973;**4**(2):73–88.

Illerbrun D, *et al.* Intervention effects with language-delayed kindergarten children: the Battlesford Language Identification and Intervention Project – Stage II. *Can J Except Child* 1985;**2**:39–42.

Ireton H, *et al.* Minnesota Preschool Inventory Identification of Children at Risk for Kindergarten Failure. *Psychol Sch* 1981;**18**(4):394–401.

Irwin JV, Huskey R, Knight N, Oltman S. A longitudinal study of the spontaneous remission of articulatory defects of 1665 school children in grades 1, 2 and 3. II: The sample. *Acta Symbolica* 1973;**5**(2):1–7.

Johnston JM, Johnston GT. Modification of consonant speech-sound articulation in young children. *J Appl Behav Anal* 1972;**5**(3):233–46.

Kosinski S. Methodological issues in the early identification of a high risk population in an urban public school system. EDRS Virginia, USA 1975; 43 Page(s); 1 Microfiche.

Lee PWH, Luk ESL, Baconshone J, Lau J, Wong V, Ko L. The Developmental Language Screening Scale: a validation and normative study. *HK J Paediatr* 1990;**1**:7–22.

Lee PWH, Luk ESL, Yu KK, Baconshone J. A developmental language screening scale for use in Hong Kong. *HK J Paediatr* 1985;**2**:152–75.

McReynolds LV. Generalization in articulation training. *Anal Intervention Dev Disab* 1981;**1**(3–4):245–58.

Manhardt JE, Hansen IM, Rescorla LA. Narrative competency outcomes of specific expressive language impairment at ages six, seven and eight. Proceedings of the Annual Boston University Conference on Language Development 1996;**20**(2):465–71.

Meier JH. Developmental Screening and Assessment. Paper presented at the National Head Start Research Conference (2nd, Washington DC, November 4–7, 1993).

Miller L, Newbill C. Innovative instructional programs for students with language or behavioural disorders. *Clin Commun Disord* 1992;**2**(2):19–31.

O'Connor C. The incidence of speech impairment among school children in N Ireland. N. Ireland: Society for the speech impaired, 1987.

Ogletree BT, Daniels DB. Communication-based assessment and intervention for prelinguistic infants and toddlers: strategies and issues. *Infants Young Child* 1993;**5**(3):22–30.

Oki J, Miyamoto A, Itoh J, Kusunoki, Y Cho, K. A longitudinal study of children with language delay at 3 years of age: later WPPSI and school attendance. *No To Hattatsu* 1992;**24**(1):3–8.

Santos LM. Intervencao em comportamento verbal de pre-escolares. *Arquivos Brasileiros de Psicologia* 1985;**37**(2):125–44.

Schwartz B. Phonetische Beschreibung ausgewahlter Sprachstorungen von Kindern und Therapiemoglichkeiten mit einer Computerartikulationshilfe (Dynamo). Berichte: Institut fur Phonetik der Universitat zu Koln;**17**:5–165.

Square PA. Treatment approaches for developmental apraxia of speech. *Clin Commun Disord* 1994;**4**(3):151–61.

Thorburn MJ, Desai P, Paul TJ, Malcolm LM, Durkin M, Davidson LL. Identification of childhood disability in Jamaica: The Ten Question Screen. *Int J Rehab Res* 1992;**15**:115–27.

Weller C. The effect of two language training approaches on syntactical skills of language deviant children. *J Learning Disab* 1979;**12**(7):470–9.

Wossum DJ. The validity of the Battelle Developmental Inventory Screening tests for early detection of developmental disorders. 1993; Lubbock, TX: Texas Technical University.

Studies not translated

Hori K, *et al.* Concerning the mother's awareness of language problems at the age of three. *Jpn J Child Psychiatr* 1976;**17**(4):236–49.

Samuel E, Lerman-Sagie T, Nevo Y, Harel S. Epidemiology of developmental disorders in children in Tel Aviv. *Harefuah* 1995;**128**(12):759–762, 823.

General bibliography

Achenbach T. Child Behaviour Checklist for Ages 2–3. Burlington, Vermont: University Associates in Psychiatry, 1992.

Alpern GD, Boll TM, Shearer MS. Developmental Profile II. Los Angeles, California: Western Psychological Services, 1984.

Alpert CL, Tiernan M, Fischer R. Milieu Language Interaction Code. Nashville, Tennessee: George Peabody College, Vanderbilt University, 1988.

Anthony A, Bogle D, Ingram TTS, McIssac MW. The Edinburgh Articulation Test. Edinburgh: Churchill Livingstone, 1971.

Aram DM, Nation JE. Preschool language disorders and subsequent language and academic difficulties. *J Commun Disord* 1980;**13**(2):159–70.

Aram AM, Morris R, Hall NE. The validity of discrepancy criteria for identifying children with developmental language disorders. *J Learning Disab* 1992;**25**(9):549–54.

Arnold KS, Myette BM, Casto G. Relationships of language intervention efficacy to certain subject characteristics in mentally retarded preschool children: a meta-analysis. *Educ Training Mentally Retarded* 1986;June:108–16.

Bain BA, Dollaghan CA. Clinical Forum: Treatment efficacy – the notion of clinically significant change. *Lang Speech Hear Serv Sch* 1991;**22**:264–70.

Baker L, Cantwell DP. A prospective psychiatric follow-up of children with speech/language disorders. *J Am Acad Child Adolesc Psychiatr* 1987;**26**:546–53.

Barnett WS, Escobar CM, Ravsten MT. Parent and clinic early intervention for children with language handicaps: a cost effectiveness analysis. *J Div Early Child* 1988;**12**(4):290–8.

Baltaxe CAM. Use of contrastive stress in normal, aphasic and autistic children. *J Speech Hear Res* 1984;**27**:97–105.

Barlow DH, Hersen M. Single-case experimental designs: strategies for studying behaviour change. 2nd ed. Boston: Allyn and Bacon, 1984.

Bates E, Beaghly M, Bretherton I, McNew S, O'Connell B, Pethick S, *et al*. Early Language Inventory. San Diego, California: San Diego State University, 1986. (Unpublished).

Bax M, Hart H, Jenkins S. Child development and child health. Oxford: Blackwell Scientific Publications, 1990.

Beitchman JH, Hood J, Rochon J, Peterson M. Empirical classifications of speech/language impairment in children: II. Behavioral characteristics. *J Am Acad Child Adolesc Psychiatr* 1989;**28**:118–23.

Beitchman JH, Wilson B, Brownlie EB, Walters H, Inglis A, Lancee W. Long-term consistency in speech/language profiles: II. Behavioral, emotional, and social outcomes. *J Am Acad Child Adolesc Psychiatr* 1996;**35**(6):815–25.

Best W, Melvin D, Williams S. The effectiveness of communication groups in day nurseries. *Eur J Disord Commun* 1993;**28**(2):187–212.

Bishop DVM. The underlying nature of specific language impairment. *J Child Psychol Psychiatr* 1992;**33**(1):3–66.

Bishop DVM, Edmundson A. Language-impaired four year olds: distinguishing transient from persistent impairment. *J Speech Hear Disord* 1987;**52**:156–73.

Bishop D, Adams C. A prospective study of the relationship between specific language impairment, phonology and reading retardation. *J Child Psychol Psychiatr* 1990;**31**(7):1027–50.

Blum-Harasty J, Rosenthal J. The prevalence of communication disorders in children: a summary and critical review. *Aust J Hum Commun Disord* 1992;**20**:63–80.

Bricker D, Squires J, Kaministi R, Mounts L. The validity, reliability and cost of a parent-completed questionnaire system to evaluate at-risk infants. *J Pediatr Psychol* 1988;**13**:55–68.

Brimer MA, Dunn LM. The English Picture Vocabulary Test. Bristol: Education Evaluation Enterprises, 1962.

Bronfenbrenner U. Is early education effective? In: Guttentag M, Struening EL, editors. Handbook of evaluation research Volume 2. London: Sage Publications, 1974.

Brown R. A first language. Cambridge, Massachusetts: Harvard University Press, 1973.

Butler J. Child health surveillance in primary care: a critical review. London: HMSO, 1989.

Bzoch K, League R. Assessing language skills in infancy. Baltimore: University Park Press, 1971.

Carrow E. Carrow Elicited Language Inventory. Austin, Texas: Learning Concepts, 1974.

Carrow E. Test for auditory comprehension of language. Austin, Texas: Urban Research Group, 1974.

Carrow-Woolfolk E. Test for auditory comprehension of language – Revised. Austin, Texas: DLM Teaching Resources, 1985.

Catts HW. Early identification of dyslexia: evidence from a follow-up study of speech-language impaired children. *Ann Dyslexia* 1991;**41**:163–77.

Cicirelli V. The impact of Head Start. Athens, Ohio: Westinghouse Learning Cooperation, 1969.

Cohen J. A power primer. *Psychol Bull* 1992;**112**(1):155–9.

Cohen J, Cohen P. Applied multiple regression/ correlation analysis for the behavioural sciences. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1975.

Cohen J, Cohen P. Applied multiple regression/ correlation analysis for the behavioural sciences. 2nd ed. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1983.

Compton A, Hutton JS. The Compton-Hutton Psychological Assessments. Haywood, California: Carousel House, 1978.

Connell PJ, Thompson CK. Flexibility of single-subject experimental designs. Part III: Using flexibility to design or modify experiments. *J Speech Hear Disord* 1986;**51**:214–25.

Conture EG. Treatment efficacy: stuttering. *J Speech Hear Res* 1996;**39**:518–26.

Cook J, Urwin S, Kelly K. Preschool language intervention – a follow-up of some within-group differences. *Child Care Health Dev* 1988;**15**:381–400.

Cook TD, Campbell DT. Quasi-experimentation: design and analysis issues for field studies. Boston: Houghton Mifflin, 1979.

Court D. Fit for the future – report of the committee on child health services. London: HMSO, 1979.

Crabtree M. Houston Test for Language Development. Chicago: Stoetting. 1963.

Crombie IK, Davies HTO. Research in health care: design, conduct and interpretation of health services research. Chichester: John Wiley and Sons, 1996.

Cronbach LJ, Furby L. How should we measure 'change' – or should we? *Psychol Bull* 1970;**74**:68–80.

Culatta B, Horn D. A program for achieving generalization of grammatical rules to spontaneous discourse. *J Speech Hear Disord* 1982;**47**:174–80.

Curtiss S, Katz W, Tallal P. Delay versus deviance in language acquisition of language impaired children. *J Speech Hear Res* 1992;**35**:373–83.

Dean EC, Howell J, Hill A, Waters D. Metaphon Resource Pack. Windsor: NFER-Nelson, 1990.

Drillien C, Drummond M. Developmental screening and the child with special needs. London: Heinemann, 1983.

Dunn LM. Peabody Picture Vocabulary Test. Circle Pines, Minnesota: American Guidance Services, 1959.

Dunn LM. Peabody Picture Vocabulary Test. Circle Pines, Minnesota: American Guidance Service, 1965.

Dunn LM, Dunn LM. Peabody Picture Vocabulary Test – Revised. Circle Pines, Minnesota: American Guidance Services, 1981.

Dunn LM, Dunn LM, Whetton C. British Picture Vocabulary Scale. Windsor: NFER-Nelson, 1982.

Eiserman WD, McCoun M, Escobar E. A cost-effectiveness analysis of two alternative program models for servicing speech disordered preschoolers. *J Early Intervention* 1990;**14**:297–317.

Enderby P, Emerson J (chapter 4). Cleft palate. In: Enderby P, Emerson J, editors. Does speech and language therapy work? London: Whurr Publishers, 1995.

Enderby P, Emerson J (chapter 5). Dysarthria. In: Enderby P, Emerson J, editors. Does speech and language therapy work? London: Whurr Publishers, 1995.

Enderby P, Emerson J (chapter 9). Voice disorders. In: Enderby P, Emerson J, editors. Does speech and language therapy work? London: Whurr Publishers, 1995.

Fenson L, Dale P, Reznick J, Thal D, Bates E, Hartung J, *et al.* Technical Manual for the MacArthur Communicative Development Inventories. San Diego, California: San Diego State University, 1991.

Ferrier E, Davis M. A lexical approach to the remediation of sound omissions. *J Speech Hear Disord* 1973;**38**:126–30.

Fey M. Language Intervention with Young Children. London: Taylor and Francis, 1986.

Fisher H, Logemann J. The Fisher-Logemann Test of Articulation Competence. Boston: Houghton Mifflin, 1971.

Frankenburg WK, Dodds JB, Fandal AW. Denver Developmental Screening Test. Denver, Colorado: University of Colorado Press, 1973.

Fudala J. Arizona Articulation Proficiency Scale. Los Angeles, California: Western Psychological Service, 1970.

Gardner MF. Expressive One-Word Picture Vocabulary Test. Novato, California: Academic Therapy Publications, 1981.

Gardner MF. Expressive One-Word Picture Vocabulary Test – Revised. Novato, California: Academic Therapy Publications, 1990.

Gauthier SV, Madison CL. Kindergarten Language Screening Test. Washington: CC Publications, 1978.

Genesee F. Bilingual language development in preschool children. In: Bishop D, Mogford K, editors. Language development in exceptional circumstances. Edinburgh: Churchill Livingstone, 1988.

Gierut JA. Maximal opposition approach to phonological treatment. *J Speech Hear Disord* 1989;**54**:9–19.

Gill G. Piagetian cognitive assessment: procedures from a variety of sources arranged for convenient use. Madison, Wisconsin: University of Wisconsin Waisman Center on Mental Retardation and Human Development. 1979 (Unpublished).

Givon T. From discourse to syntax: grammar as a processing strategy. In: Givron T, editor. Discourse and syntax Vol. 12: New York: Academic Press, 1979:81–112.

Glascoe F, Martin E, Humphrey S. A comparative review of developmental screening tests. *Pediatrics* 1990;**86**:547–54.

Goldman R, Fristoe M. Goldman-Fristoe Test of Articulation. Circle Pines, Minnesota: American Guidance Service, 1969.

Goldman R, Fristoe M, Woodcock CW. Goldman-Fristoe-Woodcock Test of Auditory Discrimination. Circle Pines, Minnesota: American Guidance Service, 1970.

Gray B, Ryan B. A language program for the non-language child. Champaign, Illinois: Research, 1973.

Griffiths R. The abilities of young children. London: Child Development Research Centre, 1970.

Guralnick MJ. Efficacy in early childhood intervention programs. In: Odom SJ, Karnes MB, editors. Early intervention for infants and children with handicaps. Baltimore, Maryland: Paul H. Brookes, 1988.

Hall P (chapter 7). Childhood articulation disorders of neurogenic origins. In: Leahy MM, editor. Disorders of communication: the science of intervention. 2nd ed. London: Whurr Publishers, 1995.

Haynes C, Naidoo S. Children with specific speech and language impairment. Oxford: Blackwell Scientific Publications, 1991.

Hedges LV, Olkin I. Statistical methods for meta-analysis. London: Academic Press, 1985.

Hedrick D, Prather E, Tobin A. Sequenced inventory of communication development. Seattle: University of Washington Press, 1975.

Hodson B. The assessment of phonological process – revised. Danville, Illinois: Interstate Printers and Publishers, Inc., 1986.

Hodson B, Paden E. Targeting intelligible speech: a phonological approach to remediation. Austin, Texas: Pro-Ed, 1983.

Ingham JC. Issues of treatment efficacy: design and experimental control. In: Olswang LB, Thompson CK, Warren SF, Minghetti NJ, editors. Treatment efficacy research in communication disorders. Rockville, Maryland: The American Speech-Language-Hearing Foundation, 1990.

Ingram D. Phonological disability in children. New York: Elsevier, 1976.

Irwig L, Tosteson ANA, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mosteller F. Guidelines for meta analyses: evaluating diagnostic tests. *Ann Intern Med* 1994;**120**:667–76.

Irwig L, Macaskill P, Glasziou P, Fahey M. Meta analytic methods for diagnostic test accuracy. *J Clin Epidemiol* 1995;**48**:119–30.

Irwin JV. The Triota: a computerised screening battery. *Acta Symbolica* 1972;**3**:26–38.

Jaeschke R, Guyatt G, Sackett DL. Users' guide to the medical literature: III. How to use an article about a diagnostic test. *JAMA* 1994;**271**:389–91.

Jowett S, Evans C. Speech and language therapy services for children Slough: NFER, 1996.

Kazdin AE (chapter 9). Statistical analyses for single-case experimental designs. In: Barlow DH, Herson M, editors. Single-case experimental designs: strategies for studying behaviour change. 2nd ed. Boston: Allyn and Bacon, 1984.

Kearns KP. Flexibility of single-subject experimental designs. Part II: Design selection and arrangement of experimental phases. *J Speech Hear Disord* 1986;**51**:204–14.

Kirk S, Kirk W. Illinois Test of Psycholinguistic Abilities (revised edition). Urbana, Illinois: University of Illinois Press, 1968.

Klackenburg G. What happens to children with retarded speech at three? *Acta Paediatr Scand* 1980;**69**(5):681–5.

Knowles W, Masidlover M. The Derbyshire Language Scheme. Derbyshire Education Authority, 1979.

Kraemer HC, Andrews G. A nonparametric technique for meta-analysis effect size calculation. *Psychol Bull* 1982;**91**(2):404–12.

Kratchowill TR. Single-subject research: strategies for evaluating change. New York: Academic Press, 1978.

Law J. Evaluating intervention for language impaired children: a review of the literature. *Eur J Disord Commun* 1997;**32**:1–14.

Lazar I, Darlington R. Lasting effects of early education: a report from the consortium for longitudinal studies. *Monogr Soc Res Child Dev* 1982;**47**(195):2–3.

Lee L. Northwestern Syntax Screening Test. Evanston, Illinois: Northwestern University Press, 1969.

Lee L. Northwestern Syntax Screening Test. Evanston, Illinois: Northwestern University Press, 1971.

Lee L. Developmental Sentence Analysis. Evanston, Illinois: Northwestern Press, 1974.

Leiter R. Leiter International Performance Scale I. Chicago: Stoelting, 1979.

Leonard L. Modelling as a clinical procedure in language training. *Lang Speech Hear Serv Sch* 1975;**6**:72–85.

Leske M. Prevalence estimates of communicative disorders in the US. *ASHA* 1981;**23**:217–25.

Lichtenstein R, Ireton H. Preschool screening: identifying young children with developmental and educational problems. New York: Grune Stratton, 1984.

Light RJ, Pillemar DB. Summing up: the science of reviewing research. Cambridge, Mass: Harvard University Press, 1984.

Lincoln MA, Onslow M. Long-term outcome of early intervention for stuttering. *Am J Speech Lang Pathol* 1997;**6**(1):51–8.

Lord FM. The measurement of growth. *Educ Psychol Measurement* 1956;**16**:421–37.

MacKeith R, Rutter M. A note on the prevalence of language disorders in young children. In: Rutter M., Martin J, editors. The child with delayed speech. London: Heinemann, 1972.

Markman E. Comprehension monitoring. In: Dickson WP, editor. Children's oral communication skills. New York: Academic Press, 1981:61–84.

Markowitz JB, Hebbeler K, Larson JC, Cooper JA, Edmister J. Using value-added analysis to examine short-term effects of early intervention. *J Early Educ* 1991;**15**(4):377–89.

McCarthy D. The McCarthy Scales of Children's Abilities. New York: The Psychological Corporation, 1972.

McCartney E, van der Gaag A. How shall we be judged? Speech and language therapists in educational settings. *Child Lang Teaching Ther* 1996;**12**:314–27.

McLean LK, Woods Cripe JW. The effectiveness of early intervention for children with communication disorders. In: Guralnick MJ, editor. The effectiveness of early intervention. Baltimore, Maryland: Paul H Brookes, 1997.

McDonald E. A deep test of articulation. Pittsburgh: Stanwix House, 1964.

McReynolds LV, Kearns KP. Single-subject experimental designs in communicative disorders. Baltimore, Maryland: University Park Press, 1983.

McReynolds LV, Thompson CK. Flexibility of single-subject experimental designs. Part I: review of the basics of single-subject designs. *J Speech Hear Disord* 1986;**51**:194–203.

Miller JF. Assessing language production in children: experimental procedures. Needham Heights: Allyn and Bacon, 1981.

Miller JF, Chapman RS. Systematic Analysis of Language Transcripts (SALT). Madison, University of Wisconsin Language Analysis Laboratory, Waisman Center on Mental Retardation and Human Development, 1985.

Miller J, Yoer D. Miller-Yoder Language Comprehension Test. Baltimore, Maryland: University Park Press, 1984.

Mogford K (chapter 7). Oral language acquisition in the prelingually deaf. In: Bishop D, Mogford K, editors. Language development in exceptional circumstances. Hove: Lawrence Erlbaum Associates, 1988.

Monahan D. Remediation of common phonological processes. Tigard, Oregon: CC Publications, 1984.

Morley M. The development and disorders of speech in childhood. Edinburgh: E&S Livingstone, 1965.

Morris Friehe M, Sanger D. Follow-up of children at risk for language problems. *J Commun Disord* 1994;**27**:241–56.

Nelson NW, Hyter YD. Black English Sentence Scoring: development and use as a tool for nonbiased assessments. Western Michigan University, Kalamazoo, 1990 (Unpublished). (Cited by NW Nelson (1993). Childhood language disorders in context: infancy through adolescence. New York: Merrill.)

Newcomer PL, Hammill DD. Test of language development – primary. Austin, Texas: Pro-Ed, 1982.

Nye C, Foster SH, Seaman D. Effectiveness of language intervention with the language/learning disabled. *J Speech Hear Res* 1987;**52**:348–57.

Olswang LB (chapter 10). Treatment efficacy: the breadth of research. In: Olswang LB, Thompson CK, Warren SF, Minghetti NJ, editors. Treatment efficacy research in communication disorders. Rockville, Maryland: American Speech-Language-Hearing Association, 1990.

Olswang LB, Bain BA. Intervention issues for toddlers with specific language impairments. *Top Lang Disord* 1991;**11**(4):69–86.

Paul R, Hernandez K, Taylor L, Johnson K. Narrative development in late talkers: early school age. *J Speech Hear Res* 1996;**39**:1295–303.

Pendergast K, Dickey S, Selmar J, Soder A. The Photo Articulation Test. Austin, Texas: Pro-Ed, 1984.

Petrie I. Characteristics and progress of a group of language disordered children with severe receptive difficulties. *Br J Disord Commun* 1975;**10**(2):123–33.

Ramey CT, Landesman Ramey S. Effective early intervention. *Mental Retardation* 1992;**30**(6):337–45.

Records NL, Tomblin JB. Clinical decision making: describing the decision rules of practising speech-language pathologists. *J Speech Hear Res* 1994;**37**:144–56.

Reid J, Millar S, Tait L, Donaldson M, Dean E, Thomson G, *et al.* Pupils with special education needs: the role of speech and language therapists. Edinburgh: SCER, 1996.

Renfrew CE. Action Picture Test. Oxford: Oxford University Press, 1986.

Reynell J. The Reynell Developmental Language Scales (revised). Windson: NFER-Nelson, 1977.

Reynell J. The Reynell Developmental Language Scales. 2nd ed. Windson: NFER-Nelson, 1985.

Roberts JE, Scheule CM Otitis Media and later academic performance: the linkage and implications for intervention. *Top Lang Disord* 1990;**11**(1):43–62.

Robson C. Real world research: a resource for social scientists and practitioner-researchers. Oxford: Blackwell, 1993.

Rosenthal R (chapter 16). Parametric measures of effect size. In: Cooper H, Hedges LV, editors. Handbook of research synthesis. London: Russell Sage Foundation, 1994.

Russell VJ (chapter 10). Cleft palate and other craniofacial anomalies in children. In: Leahy MM, editor. Disorders of communication: the science of intervention. 2nd ed. London: Whurr Publishers, 1995.

Ryan B, Baker RD. Programmed conditioning for articulation. Monterey, California: Monterey Learning Systems, 1971.

Salzberg CL, Strain PS, Baer DM. Meta-analysis for single-subject research: when does it clarify, when does it obscure. *Remedial Spec Educ* 1987;**8**(2):43–8.

Schweinhart LJ, Weikart DP. A summary of significant benefits: The High/Scope Perry Preschool Study through age 27. Ypsilanti: High/Scope, 1993.

Scollon R. Conversations with a one year old: a case study of the developmental foundation of syntax. Honolulu, Hawaii: The University Press of Hawaii, 1976.

Scruggs TE, Mastropieri MA, Casto G. The quantitative synthesis of single-subject research: methodology and validation. *Remedial Spec Educ* 1987;**8**(2):24–33.

Scruggs TE, Mastropieri MA, Forness SR, Kavale KA. Early language intervention: a quantitative synthesis of single-subject research. *J Spec Educ* 1988;**22**(3):259–83.

Sheehan R, Gallagher RJ. Methodological concerns in evaluating early intervention. *Diagnostique* 1983;**8**:75–87.

Sheridan MD, Peckham C. Follow-up at 11 years of children who had marked speech defects at 7 years. *Child Care Health Dev* 1975;**1**:157–66.

Shonkoff JP, Hauser-Cram P. Early intervention for disabled infants and their families: a quantitative analysis. *Pediatrics* 1987;**80**(5):650–8.

Shriberg LD. Developmental phonological disorders: moving towards the 21st century-forwards, backwards or endlessly sideways? *Am J Speech Lang Pathol* 1994;**3**:26–8.

Shriberg LD, Kwiatkowski J. Phonological disorders III: a procedure for assessing severity of involvement. *J Speech Hear Disord* 1982;**47**:256–70.

Sidman M. Tactics of scientific research: evaluating experimental data in psychology. New York: Basic Books, 1960.

Sloane P (chapter 8). Voice disorders in children. In: Leahy MM, editor. Disorders of communication: the science of intervention. 2nd ed. London: Whurr Publishers, 1995.

Sommers RK. Articulation disorders. Englewood Cliffs, New Jersey: Prentice Hall, 1983.

Sparrow S, Balla D, Cicchetti D. Vineland Adaptive Behaviour Scales: Interview Edition. Circle Pines, Minnesota: American Guidance Service, 1984.

Stark RE, Bernstein LE, Condino R, Bender M, Tallal P, Catts H. Four-year follow-up study of language impaired children. *Ann Dyslexia* 1984;**34**:49–68.

Stewart-Brown S, Farmer A. Screening could seriously damage your health. *BMJ* 1997;**314**:533–4.

Stevenson J. Developmental changes in the mechanisms linking language disabilities and behavior disorders. In: Beitchman JH, Cohen NJ, Konstantareas MM, Tannock R, editors. Language, learning and behavior disorders: developmental, biological and clinical perspectives. Cambridge: Cambridge University Press, 1996.

Studdert-Kennedy M, Mody M. Auditory temporal perception deficits in the reading-impaired: a critical review of the literature. *Psychonomic Bull Rev* 1995;**2**(4):508–14.

Sturner R, Layton T, Evans A, Heller J, Funck S, Machton M. Preschool speech and language screening: a review of currently available tests. *Am J Speech Lang Pathol* 1994;January:25–36.

Tallal P, Piercy M. Developmental apasia: impaired rate of non verbal processing as a function of sensory modality. *Neuropsychologia* 1973;**11**:389–98.

Tallal P, Allard L, Miller S, Curtiss S (chapter 10). Academic outcomes of language impaired children. In: Hulme C, Snowling M, editors. Dyslexia: biology, cognition and intervention. London: Whurr, 1997.

Templin M, Darley F. Templin-Darley Tests of Articulation, 2nd ed. Iowa City, Iowa: University of Iowa, 1969.

Templin MC. Certain language skills in children: their development and interrelationships. Child Welfare Monographs, 26. Minneapolis, Minnesota: University of Minnesota Press, 1957.

The Department for Education. Code of practice for special educational needs. London: HMSO, 1994.

The Scottish Office: Education and Industry Department. The education of pupils with language and communication disorders. Edinburgh: HMSO, 1996.

Thorndike RL, Hagen EP, Sattler JM. Stanford-Binet Intelligence Scale. 4th ed. Chicago: Riverside Publishing Company, 1986.

Tizard B, Hughes M. Young children learning. London: Fontana, 1984.

Uzguris I, Hunt J McV. Assessment in infancy: ordinal scales of psychological development. Urbana, Illinois: University of Illinois Press, 1975.

Urwin S, Cook J, Kelly K. Preschool language intervention: a follow-up study. *Child Care Health Dev* 1988;**14**:127–46.

Warnock M. Special educational needs. Report to the committee of enquiry into the education of Handicapped children and young people. London: HMSO, 1978.

Wells G. Language development in the pre-school years. Cambridge: Cambridge University Press, 1985.

Weiner F. Phonological Process Analysis. Baltimore, Maryland: University Park Press, 1979.

Weiner F. Treatment of phonological disability using the method of meaningful minimal contrast: two case studies. *J Speech Hear Disord* 1981;**46**:97–103.

Werner ED, Kresheck JD. Structured Photographic Expressive Language Test-II. Illinois: Janelle Publications, 1974.

Werner ED, Kresheck JD. Structured Photographic Expressive Language Test – Preschool. Illinois: Janelle Publications, 1983.

White OR. Some comments concerning 'The quantitative synthesis of single-subject research'. *Remedial Spec Educ* 1987;**8**(2):34–9.

Whitehurst GJ, Fischel JE. Practitioner review: early developmental language delay: what, if anything, should the clinicians do about it? *J Child Psychol Psychiatry* 1994;**35**(4):613–48.

Wilson JM, Jungner G. Principles and practice of screening for disease. WHO public health papers no. 34. Geneva: WHO, 1968.

Wright BA, Lombardino LJ, King WM, Puranik CS, Leonard CM, Merzenich MM. Deficits in auditory temporal and spectral resolution in language-impaired children. *Nature* 1997;**387**:176–8.

Zachman L, Huisingh R, Jorgenson C, Barrett M. Language Sentence Imitation Screening Test. Illinois: Linguisystems, 1982.

Zigler E, Muenchow S. Head Start: the inside story of America's most successful educational experiment. New York: Basic Books, 1992.

Zimmerman IL, Steiner VG, Evatt R. The Preschool Language Scale. Columbus, Ohio: Charles E. Merrill, 1969.

Zimmerman IL, Steiner VG, Pond RE. The Preschool Language Scale. 2nd ed. San Antonio, Texas: Psychological Corporation, 1979.

Zimmerman IL, Steiner VG, Pond RE. The Preschool Language Scale – 3. San Antonio, Texas: Psychological Corporation, 1992.

Zwart EW. Effectiveness of language intervention with language impaired children: a literature review. The Institute of Phoniatrics, Utrecht University. 1996 (Unpublished).

Appendix 1

Case descriptions for delayed speech and language development

Child 1: Expressive language delay – persistent

David is the second of two children in a professional family with no family history of speech and language delays. Apart from brief periods of ear infection for which he received antibiotics, he had no significant medical history. He was a communicative baby, babbling by 1 year. By 18 months he relied entirely on pointing to have his needs met and this practice went on until his first word at 23 months. His vocabulary developed very slowly and he tended to use gesture accompanied by vowel sounds. It seemed fairly clear by 30 months that he wanted to say more than he was able to and his frustration often ended in tantrums. He started to use two-word utterances "more ball" and "mummy car" by 3 years, and was stringing together short telegrammatic sentences by the time he went into nursery at $3\frac{1}{2}$. He seemed to have a reasonable level of single-word vocabulary at this stage but word combination and particularly modifying words to mark tense, numbers, etc., were very difficult for him. Despite his single-word output he tended to use the same very limited range of verbs, notably **get** and **do** at every opportunity. His parents reported that he was able to understand what was said to him, and assessment by a speech and language therapist indicated that his comprehension was indeed within normal limits.

In nursery his language developed but he proved quite difficult to understand because his speech seemed very muffled. He appeared very self-conscious about speaking and tended to hold back in his peer group rather than commit himself to speaking in front of them at the request of his teacher. The initial response of his reception class teacher was to say that he was shy rather than delayed in his language development. However, his confidence developed through the year but he remained relatively monosyllabic and tended to express himself in boisterous games with the other boys rather than trying to respond verbally.

Although relatively slow in acquiring literacy skills he moved from whole-word reading to the use of phonics by about the age of 7 years, much later than most other children in his class. His writing was better than his reading and there was some indication that he preferred writing to speaking as a means of expression in class. Nevertheless he continued to exhibit errors long after they had disappeared in the work of most of his peers. At 7 years he was still cause for concern to his teachers. He did not have a statement of educational need but it was widely recognised that he was only able to perform appropriately if given plenty of time to formulate sentences. There was considerable concern that he would do very badly on the standard assessment tasks for speaking and listening tasks in the National Curriculum.

Child 2: Expressive language delay – transient

George is the first child of a family with no history of speech or language delays. His parents were very busy people when he was a baby, and although they spent time with him, they freely admitted that they did not know what to do to help him because they had no experience of other young children.

He achieved his motor milestones on time but showed little interest in communication except for pointing at familiar objects to draw them to the attention of close family members. Similarly he showed little interest in adults or children outside the immediate family. His mother became concerned about his development when he went into a work place crèche at 26 months and she began to make comparisons with other children.

Although he was still very 'clingy' with his parents in their presence, he settled well into the nursery and began trying to communicate with the other children and with staff members. Initially he appeared to have difficulty putting words together but at 27 months he began to use two-word utterances with one member of staff in the crèche and obviously enjoyed the response he received.

By 4 years of age he was formulating sentences, and although they were often rather simpler sentences that those of many of his peers, they were well

formed and his vocabulary was developing well. At school entry his parents felt sufficiently confident about his development not to draw it to the attention of his new school. Despite their apprehension he adapted well to his new environment and quickly made friends. Like many children of 5 he remained reluctant to speak in the group but his teacher made a special effort to make him more confident and by the age of 6 he was indistinguishable from his peers in all his academic subjects, including his reading and writing.

Child 3: Speech delay – persistent

Joe was difficult to feed as a baby but his parents put it down to his having been born 6 weeks prematurely and to his just being a 'difficult' child. Like most babies he was a very messy feeder but this continued into his second year. His father was reported to have been hard to understand as a child and although he speaks clearly now, he remembered struggling with his literacy skills as a child, and as an adult rarely reads for pleasure.

By 2 years Joe made noises and clearly was trying to express himself, modifying his vowel sounds but tending to retain a single /d/ sound in the place of all initial consonants, and omitting final consonants. At this stage he was completely unintelligible to all but his closest family. By his third birthday he continued to be very difficult to understand and still tended to dribble when he was tired. His expressive language was very limited even though he often attempted to communicate. By contrast, his comprehension was reported to be within normal limits. His parents commented on his clumsiness, though no referral was made to physiotherapy.

Other children found him difficult to understand in nursery. Although this did not seem to concern them when he started in nursery, by the time he reached school age they had become very aware of his difficulty and would comment on it to the teacher. At 4 years he suddenly became very self conscious and refused to try to speak. His speech did improve, particularly following a period in speech and language therapy. This consisted of his practising sounds, which he found very difficult, and listening to differences between sounds. In fact, he was usually able to discriminate between sounds particularly if they were presented to him in single words. But when he tried to formulate his own sentences he had to struggle to convey his message.

In school he also struggled with reading. He found it very difficult to progress from the single-word reading stage to developing his phonic skills. By 7 he had become clearer in his speech, particularly when talking about familiar topics, and he was making progress with his literacy skills. After the 7-year standard assessment tasks he was given a statement of educational need and provided with regular speech and language therapy. By the age of 9 his speech difficulties had largely resolved, though his speech sometimes became less intelligible if he was tired or stressed. His literacy skills remained problematic for many years to come.

Child 4: Speech delay – transient

Jean's birth was normal and her early years were without medical complications except for the usual round of coughs and colds and a mild fluctuating hearing loss. She was a few months slower at starting to speak relative to other children in the crèche, which she attended from her first birthday. Her mother was not particularly worried about this because her older sister had also been a little slow and it had not made any difference to her progress. By $2^{1}/_{2}$ years she had started to put two words together but it was impossible for anyone apart from her mother to understand what she was saying. Her mother admitted that out of context she often found it hard to decipher Jean's intended meaning. She was referred by her health visitor to a speech and language therapist who noted phonological delay characterised by a highly systematic set of errors – in particular, she had no evidence of fricatives (/f/s/sh/) and only two plosives (/d/b/). Her vowel system was intact. At this assessment her receptive language was reported to be within normal limits. By the 6-month review, she had started to use some fricatives in word-initial position and now had a reasonable range of plosives. She was intelligible to immediate family, but strangers still found her very difficult to understand. Assessment of her discrimination of speech sounds indicated that she could perceive the differences between the sounds that she still was not able to produce. At this stage (3;9 years) she was in nursery and had not started to become self conscious about her speech. She received some speech and language therapy for a limited period. The work focused on her production of sounds. By 4;6 she was able to produce most of the speech sounds in single words but was still having difficulty producing all of them correctly. This meant that she was sometimes reluctant to volunteer explanation in class. She was recognising familiar words and had made some progress into the orthographic

stage of literacy development. Her teacher noted that Jean was of a comparable level to all the other children in the class.

Child 5: Expressive/receptive language delay – persistent

Natalie's birth was difficult. She was born at full term but spent a week in intensive care. She was discharged and not followed-up by her local child development centre. Her mother reported that she was a quiet baby, something which she welcomed at the time. There was a family history of slow language development. Natalie passed her 8-month developmental check but when she went back at 18 months her health visitor felt that she ought to be communicating more, and after reviewing her 3 months later, referred her for speech and language therapy. Assessment indicated that her parents found her to be a difficult child to communicate with, which resulted in their leaving her to her own devices. She had a toy doll which she enjoyed carrying around but she showed little evidence of exploratory representational and later symbolic play. She turned to her name but found it difficult to listen to what was said to her, tending to flit from situation to situation without commenting on what she found. She was referred for audiology and was found to have normal hearing and middle-ear pressure. She went into nursery at 3 years and observations revealed that she continued to spend a considerable amount of her time moving from one activity to the next and interacted relatively little with the other children. She was referred to her local child development centre at 3;3 years where she was found to have a developmental quotient within normal limits, albeit in the low average range. Her behaviour was described by her parents as 'difficult'. She was prone to tantrums of frustration and went through periods of soiling. Occupational therapy and physiotherapy reports suggested that she was having difficulty with some hand–eye coordination tasks and was not generally very well coordinated in her gross motor skills. In both cases the level of difficulty was not sufficient to warrant intervention.

She attended for relatively brief speech and language therapy groups between 3 and 4 years and, although these often helped her parents interact more effectively with her and improved her listening and comprehension skills, she continued to have expressive and receptive language scores ranging between –1.7 and –2.3 SDs below the mean. In addition it was noted that her speech was not as clear as that of most of her peers and she had a tendency to stammer when under any sort of pressure. By $4\frac{1}{2}$ she was put forward for a statement of educational need and was admitted to a language unit integrated within a mainstream school shortly afterwards. She responded well to the highly structured day of the unit because it seemed to enable her to predict more of what was expected within school. Her concentration began to become more integrated and she became more compliant within the class. Her comprehension on standardised testing improved somewhat but the standard score remained well outside normal limits. Her speech improved but her ability to convey concepts remained very limited. The more abstract the language required of her the more apparent her difficulties became. For example, while she could talk about a picture placed in front of her she found it very difficult to express temporal concepts, in part because she lacked the sequencing abilities, but also because she could not mark the necessary changes to the verbs concerned.

Literacy presented a range of problems for her. Indeed by 7 years of age she could do no more than recognise a handful of words. The only strategy she had for dealing with unfamiliar words was to identify the first letter and then search for a word of equivalent length which started with that letter. Natalie continued to experience difficulties related to her language but these difficulties were often construed rather differently by her school teachers. She struggled with all areas covered by the standard assessment tasks at 7 years, most notably her speaking and listening skills, her literacy and her maths work. She found it hard to relate to many of her peers, preferring to spend time with children in the nursery, presumably because they had equivalent levels of communication skill. Many teachers expressed concerns that she was developmentally delayed. Full developmental review indicated that she continued to have disproportionate difficulties in her language relative to her non-verbal skills.

Child 6: Expressive/receptive language delay with associated pragmatic difficulties – persistent

Miles was the last of a family of five children. He was born at full term and presented with no medical difficulties except for febrile convulsions in his first 3 years. Looking back his parents described him as an irritable and fractious baby. However, nothing was done about this at the time. By 18 months he actively seemed to resist adult attentions, though he would point to indicate

needs. He used his first words at $3^{1}/_{2}$ years: by this time his parents had expressed concern and had had him assessed in his local paediatric assessment centre. The assessment signalled a primary language delay with both expression and comprehension affected. It was noted that he spent a considerable time watching television and became very excited when favourite advertisements appeared on the screen. He exhibited very little symbolic play. Although he recognised the function of most toys he would not choose to play with them and was not apparently able to transform them in a symbolic manner. He enjoyed familiar routines and became anxious if they were altered for any reason. He continued to present as an awkward child preferring his own company to that of his peers much of the time, though by 4 he was relating to adults. Standardised language assessment at 4 years indicated receptive and expressive scores both more than 2 SDs below the mean. A common observation in the notes indicated a recognition that his difficulties were not confined to his language skills and that there was an unusual quality to his interaction, suggesting high levels of anxiety associated with the process. He initiated interaction infrequently and when he did, it was often unsuccessful. On these occasions he seemed to have no way of renegotiating with his peers to get the conversation going again. He continued to show a considerable interest in the television and was by now repeating familiar passages.

Although he remained quite withdrawn when he started in school, his teacher noted his apparently advanced memory of familiar topics and his highly developed reading skills. He also demonstrated an ability to manipulate numbers. On testing he showed a difficulty reading for meaning but his structural literacy skills meant that he was in advance of most children in his class at the beginning of the school year. This advantage had largely disappeared by the end of the year but he remained relatively able in this respect. When formally tested at 6 years of age his comprehension seemed to have improved considerably. He responded well to the highly structured format of the testing environment. Indeed his performance in these circumstances was much better than would be anticipated given his class performance. He stood out in any social activity as being completely unable to understand what his peers were talking about once they started to joke or make use of complex inference. This pattern of a positive response to highly decontextualised language relative to his difficulties with language in context remained one of the characteristics of the profile of his communication skills. In terms of his other classroom performances he continued to perform very erratically, apparently enjoying tasks to which he could apply rote learning strategies but finding it very difficult to discuss the results. While this difficulty was less apparent in his first year in school, by year 3 it had begun to separate him out from other children. Inevitably this presented him with disproportionate difficulties in the speaking and listening tasks of the National Curriculum. Of greater concern was his continuing difficulty in forming friendships.

Appendix 2

Intervention descriptions

The literature on intervention for children with speech and language delays covers four broad areas:

- didactic approaches
- naturalistic approaches
- hybrid approaches
- other approaches.

These reflect common practice, though speech and language therapists and educationalists do not necessarily explicitly conceptualise their work in this way. For example, they may draw upon more than one approach at any one time and may make client-focused modifications as the interventions progress. These approaches provide a broad framework but do not in themselves presuppose a rationale for the target behaviours selected. Similarly they do not presuppose the location of therapy which may take place in the home, the clinic or the educational environment.

Didactic approaches

Terms that are often associated with this approach are **elicited imitation** or **mand modelling**. In essence, the child is given a model of a sound, a word, a communication behaviour or a syntactic construction, and a direct attempt is made to elicit the child's production of that model. Usually this happens after a number of presentations of the model. Programmes vary as to the amount of time they allocate to input activities, or activities devoted to providing the child with the correct form of the utterance. Similarly there are differences with the methods used to elicit the required form. In some cases the therapist will use positive reinforcement of a verbal nature, in others a reward system based on an immediate or deferred reward will be used.

This technique has been widely used for the teaching of vocabulary items and speech sounds, though over the years the emphasis has shifted from the elicitation to the modelling stage. It has also been used in the teaching of syntax and skills related to the social use of language (e.g. turn taking or conversational initiation).

Advocates of this approach suggest that it is successful in providing a quick therapeutic response when the targets are transparent. Whether or not it is possible to demonstrate generalisation of the acquired behaviours has been the subject of some investigations. It has been questioned by those who believe that the child is inherently programmed to identify language rules. It is seen as teaching specific behaviours, not encouraging the child to make the links for themselves, and promoting the learning of language in a social vacuum. The argument runs that children are not normally taught language in such an explicit fashion and that it does not, therefore, make sense to adopt this technique in intervention. By contrast, advocates suggest that the children concerned are not making these links because the nature of their language difficulties stops them from doing so and that formal training of this type is the only way forward.

In practice, elicited imitation is a constituent part of most intervention programmes because a child's expression is often the only way to measure change following intervention.

Naturalistic approaches

Terms which have been used to describe this approach are **interactionist or incidental teaching**. This group of therapies emphasises the need for generalisation of any language gains and seeks to emulate the context in which language is learned in the normal process of language acquisition. They are widely used with children in the very early stages of language development, though it has also been adopted in the later stages of the therapeutic process once specific skills have been trained. For example, it has been used as a means of promoting generalisation once specific sounds have been taught.

Therapy is aimed at promoting functional language and frequently involves parents as active participants in the process. There are a series of underlying behaviours adopted by the adults, most notably 'contingent responsiveness', by which the parent or therapist responds to the focus of attention of the child rather than imposing a

different context specifically geared to therapeutic ends. One programme that clearly articulates the aims of the naturalistic approach is the Hanen Early Language Parent Programme which speaks of the '3 As': **A**llowing the child to lead, **A**dapting to the child's needs, and **A**dding language at an appropriate level to that of the child. For example if the child is at a single-word level, the adult will model the intended response adding another relevant word. The child says "**Bus**" and the adult says "Yes, that's a **big bus**" or "That's **your bus**". The involvement of the parent encourages the use of these techniques in the home as well as in the therapeutic/educational context.

Most therapeutic strategies involved in naturalistic approaches to intervention are general in nature and there is some question as to how effective the training is in promoting specific linguistic behaviours. However, if one assumes that this approach is promoting the development of the child's existing language functioning rather than specifically teaching new words it is firmly rooted in normal development.

Hybrid approaches

Hybrid approaches are a combination of didactic and naturalistic approaches. A term commonly used to describe hybrid approaches is **milieu therapy**. This refers to intervention programmes that explicitly draw on both didactic and naturalistic techniques. In some cases these are explicitly listed as a part of the therapy programme and commonly include mand modelling, incidental teaching, and time-delay techniques (when a controlled response period is inserted to encourage the child to respond). A typical example might be intervention with a child targeting the comprehension of two-word constructions using a verb and an object. The therapist/teacher might set up the context such that a child (either on his own or with other children) is presented with a series of objects. The function of each object is modelled perhaps five times (you **kick** the **ball**). The set of objects is introduced on another occasion perhaps with some other children and the time delay introduced (What do you do with this? You..... **kick ball**. The teacher waits for a response and after an allotted time models again if the child has not produced it). All the other adults in the child's environment model the **verb + object** construction to the child where appropriate. Once it has been established it is then appropriate to incorporate the mand approach in which the teacher explicitly asks the child to repeat the construction. Finally, the effect is measured in terms of the extent to which the construction is generalised to other situations. For most children a number of such goals would be monitored simultaneously.

Although this approach is particularly clearly articulated in intervention studies from the USA it probably mirrors much of the work in 'language units' in the UK. These are units set up to provide structured input to children with primary language delays. They are usually placed within a mainstream school and the children benefit from both the specialised structured input and from the opportunity to integrate into the mainstream class with their peers.

Other approaches

Teachers and therapists are constantly exploring different ways to promote language development and this results in the development of new techniques some more and some less naturalistic. To indicate something of the range four are illustrated below ranging from the highly context dependent to the largely context free.

Non-directive therapy

This approach is largely intended for children who are non-communicating or for whom the experience of communication is perceived as threatening. The child chooses a topic usually from a predetermined set of materials, and chooses how to use the space in the room. The therapist then sits at an appropriate distance from the child and provides a running commentary on whatever it is that the child is choosing to do with the materials. The purpose of such an approach is to give the child a favourable experience of communication and to promote initiation. While it might involve the definition of interaction targets the outcomes for such an approach tend to be general in nature.

Auditory training

This approach assumes that the underlying difficulty for the child is attention to the auditory medium. This may be as a function of intermittent hearing loss or a more specific inability to process sounds or a general inattention associated perhaps with hyperactivity. The teacher/therapist encourages the child to listen and look at the source of a sound, to discriminate its meaning and to respond accordingly. There is often a strong behavioural component in the intervention. The child is rewarded as a direct response to the improved listening skills. Auditory training is widely used

as a part of intervention with most children with speech and language delays. It has a more specific application when applied to listening to and discrimination of specific phonemes. This latter application is common as a component of interventions for speech delays.

Comprehension monitoring

This is a multi-stage approach designed to promote the attention and comprehension skills of language-delayed children. It focuses on providing the children with appropriate strategies for coping with conversations. Good listening behaviours are modelled and the child is encouraged to reflect on what makes for good listening and the effect that it has on an interaction and on the way they feel about it. Children role play good and poor listening behaviours to establish control over their own communicative behaviours. Children are then given direct and indirect models of appropriate ways of eliciting information from others and of ways of retaining requests for long enough to process them. Particular requests are given to the child in which insufficient information is provided for the child to respond appropriately. The child is then encouraged to reflect upon what it was that made it difficult for him to understand. The situation is then manipulated so that, for example, a child has to ask for further information from a teacher who mumbles or has to ask the meaning of unfamiliar but crucial words in a story.

Cognitive therapy

At any given time therapists and teachers are involved with both the behavioural and the cognitive aspects of intervention. For many children it may be necessary to obtain a clear insight into how the child processes language. Specific processing difficulties may be remediated by asking the child to reflect on specific linguistic structures. Widely used with older language delayed children and with adults with acquired language disorders, this approach often involves presenting structured materials to the client, and modelling an appropriate response such as selecting pairs of similar sounding words or words which are semantically similar. Thereafter the child is asked to make judgements about the appropriacy of certain combinations of words ("Which ones go together?") or the grammaticality of specific sentences. The logic behind this approach is that, although largely free of the context it is highly specific in the linguistic domains that it is able to tap. Furthermore, the emphasis on internal processes rather than external behaviours is intended to show that the child has the prerequisite linguistic knowledge to communicate effectively. In essence it is intended to access the child's thought processes rather than the more obviously behavioural aspects of communication.

Appendix 3

Sources of information and search strategies

Database searches

Cochrane library database search

Cochrane Database of Systematic Reviews (CDSR)
Database of Abstracts of Reviews of Effectiveness (DARE)
The Cochrane Controlled Trials Register (CCTR)
The Cochrane Review Methodology Database (CRMD).

Terms searched

- speech therapy
- language therapy
- speech delay
- speech screening
- language delay
- language screening
- language & child & therapy
- child voice
- child language disorder
- child speech disorder
- speech delay identification
- language delay identification

(All terms searched with AND operator as no phrase searching available.)

Result

For all terms except 'speech therapy': no references found or no useful references for required age range. For the term 'speech therapy' three references were found.

Decision process for selecting literature databases

CROS and DialIndex search across databases

Both CROS and DialIndex search mechanisms were used, as their coverage of databases is not the same.

Sections searched

CROS: 80 (Social Sciences and General reference).
DialIndex: EDUCAT, LANGUAGE, MAGTEXT, MEDTEXT, SOCSCI, MEDICINE, PSYCH
(This gave access to 60 DialIndex files).

Terms searched

A. (speech or language) and screening
B. language and children and therapy

Analysis

The searches produced rankings of databases.

The following databases featured in the top 10 rankings but were excluded because they were searched on a full text basis, which did not compare with the non full text searches: Reuter Textline: 80–88; 89–92; 93–; IAC Magazine Database 59–.

The following databases were excluded by inspection, their emphasis being away from our field: IAC Business ARTS; IAC (SM) Health & Wellness DB(SM); CAB Health (= HUMN); Health star; Management Contents (= MGNT); AMA Journals; Asia/Pacific Business Journals; Mental Health Abstracts.

Each term (A and B) yielded two rankings, one each from CROS and DialIndex (see below).

A. (speech or language) and screening

CROS	DialIndex
1. ERIC 1966	1. ERIC 1966
2. Psychological INFO Psychological Abstracts	2. EMBASE 1974
3. Current Contents 1992	3. MEDLINE 1966
4. Dissertation Abstracts 1861	4. Psychological INFO 1967
5. Social Scisearch 1972	5. Dissertation Abstracts
6. CINAHL 1983	6. Biosis Previews 1969

B. language and children and therapy

CROS	DialIndex
1. Psychological INFO Pyschological Abstracts	1. LLBA
2. ERIC	2. EMBASE 1974
3. Current Contents	3. MEDLINE 1966
4. CINAHL 1983	4. Psychological INFO 1967
5. Social Scisearch (= BIDS)	5. ERIC 1966
6. Dissertation Abstracts	6. Scisearch

Conclusion
Databases selected for the review:

- EMBASE
- MEDLINE
- ERIC
- PsychINFO (= PsychLit)
- CINAHL
- LLBA

This assumes that the coverage of Current Contents is within the other databases selected.

Unpublished literature: database searches

Two databases dealing with unpublished literature were scanned for any relevant material.

System for Indexing Grey Literature in Europe (SIGLE)

Terms searched
- speech therapy
- (screening-methods) AND speech
- speech disorders
- screening AND speech
- language disorders

Result
No relevant records found.

BOSTON SPA CONFERENCES (British Library database)

Terms searched
- speech disorders
- language disorders
- speech AND screen$
- speech therapy
- language therapy
- early intervention
- therapy efficacy
- therapy outcome

Result
Seven references found for consideration.

Search strategies for literature databases

Details of search terms for the following databases are presented below: CINAHL, EMBASE, ERIC, LLBA, MEDLINE, PsychLit.

CINAHL strategy
OVID 1982–November 1996.

/ denotes thesaurus term entry
.tw denotes free text search

Set	Search	Results
001	language disorders/	152
002	exp speech disorders/	410
003	voice disorders/	50
004	communicative disorders/	110
005	"Verbal Impairment (Saba HHCC)"/	0
006	"Impaired Verbal Communication (NANDA)"/	11
007	or/1-6	672
008	(speech or language or voice or verbal).tw	2486
009	(impair$ or delay$ or problem$ or disorder$ or retard$ or disabilit$ or difficult$ or handicap$).tw	22699
010	8 adj5 9	365
011	7 or 10	865
012	limit 11 to (infant or preschool child or child)	280
013	11 and (child$ or baby or babies or infant$ or toddler$ or preschool$)	225
014	12 or 13	294
015	morbidity/	389
016	prospective studies/	4419
017	epidemiology/	517
018	research methodology/	2393
019	cross sectional studies/	1346
020	(prevalen$ or incidence or morbidity).tw	4606
021	or/15-20	12434
022	21 and 14	28
023	child development disorders/	258
024	health screening/	1233
025	health promotion/	2567
026	language tests/	140
027	language/ev	43
028	clinical assessment tools/	2468
029	nursing assessment/	2838
030	"referral and consultation"/	881
031	diagnosis, developmental/	211
032	risk factors/	4874
033	denver developmental screening test.tw	37
034	(reynell or fluharty).tw	4
035	development inventory.tw	19
036	(parent$ adj3 (checklist$ or survey$ or questionnaire$)).tw	162
037	(screen$ or test$ or diagnos$ or predictive or predictor$ or detect$ or surveillance).tw	23899
038	or/23-37	35072

039	validation studies/	894
040	exp predictive validity/	147
041	"reliability and validity"/	1804
042	research methodology/	2393
043	instrument validation/	827
044	correlation coefficient/	1321
045	(predictive or concurrent$ or valid$ or reliab$ or standardi$ or sensitivity or specificity).tw	6967
046	or/39-45	11363
047	38 and 46	5052
048	47 and 14	22
049	language therapy/	21
050	speech therapy/	135
051	speech-language pathology/	79
052	"rehabilitation, speech and language"/	105
053	(therap$ or treat$ or interven$ or rehabilit$ or train$ or educat$ or teach$ or program$ or facilitat$).tw	67294
054	(imitation or stimulation or milieu).tw	1055
055	parent child interaction.tw	17
056	or/49-55	67958
057	random assignment/	849
058	(random$ adj5 (allocat$ or assign$ or control$)).tw	1230
059	((singl$ or doubl$ or trebl$ or tripl$) adj5 (blind$ or mask$)).tw	293
060	exp clinical trials/	1998
061	(clinical$ adj5 trial$).tw	739
062	(placebo$ or random$).tw	3478
063	clinical research/	1621
064	control group/	199
065	nonrandomized trials/	13
066	therapeutic trials/	4
067	experimental studies/	1059
068	program evaluation/	1757
069	evaluation/	221
070	evaluation research/	1897
071	comparative studies/	3963
072	exp pretest posttest design/	816
073	(control$ or prospectiv$ or volunteer$).tw	12774
074	(compar$ adj5 (report$ or stud$ or trial$)).tw	2586
075	exp quasi-experimental studies/	715
076	(multiple baseline$ or time series or quasi?experimental).tw	122
077	or/57-76	25725
078	77 and 56	12200
079	78 and 14	38
080	"outcomes (health care)"/	2800
081	exp prospective studies/	4454
082	cross sectional studies/	1346
083	prognosis/	590
084	research methodology/	2393
085	follow?up.tw.	214
086	(natural adj3 history).tw	123
087	(prognosis or outlook or future or prospective).tw	8050
088	or/80-87	17801
089	88 and 14	29
090	*down syndrome/	166
091	*mental retardation/	686
092	*mental disorders/	1052
093	*cerebral palsy/	306
094	*hearing disorders/	374
095	*autism/	93
096	or/90-93	2168
097	or/91-95	2474
098	review.pt.	7676
099	meta analysis/	194
100	review literature/	0
101	meta?analy$.tw.	1
102	(systematic$ adj4 (review$ or overview$)).tw	61
103	or/98-102	7867
104	103 and 14 and 38	13
105	103 and 14 and 56	17
106	22 not 96	28
107	48 not (96 or 106)	18
108	79 not (97 or 106 or 107)	24
109	89 not (97 or 106 or 107 or 108)	4
110	104 not (96 or 106 or 107 or 108 or 109)	8
111	105 not (97 or 106 or 107 or 108 or 109 or 110)	9

EMBASE strategy

TELNET BIDS 1980–December 1996.

Upper case represent searching on EMTREE terms
Lower case represent searching text across title, keywords and abstract

1. exp LANGUAGE DISABILITY
2. exp SPEECH DISORDERS
3. VOICE DISORDERS
4. LANGUAGE DEVELOPMENT DISORDERS
5. speech development
6. communication disorder
7. speech impair*, speech delay*, speech problem*
8. speech disorder*, speech retard*, speech disabilit*
9. speech difficult*, speech handicap*
10. language impair*, language delay*, language problem*
11. language disorder*, language retard*, language disabilit*
12. language difficult*, language handicap*
13. voice impair*, voice problem*
14. voice disabilit*
15. voice difficult*, voice handicap*

16. 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15
17. exp CHILD
18. child*, baby , babies
19. infant* , toddler* , preschool*
20. 17,18,19
21. 20 AND 16
22. DOWN SYNDROME
    search as major thesaurus term
23. MENTAL RETARDATION
    search as major thesaurus term
24. CEREBRAL PALSY
    search as major thesaurus term
25. MENTAL DISORDERS
    search as major thesaurus term
26. HEARING DISORDERS
    search as major thesaurus term
27. AUTISM
    search as major thesaurus term
28. 22,23,24,25
29. 23,24,25,26,27
30. exp MORBIDITY
31. EPIDEMIOLOGY
32. prevalen*, incidence , morbidity
33. 30,31,32
34. 33 AND 21
35. 34 NOT 28 **prevalence set**
36. exp SCREENING
37. LANGUAGE TEST
38. POPULATION SURVEILLANCE
39. screen*, test*, diagnos*
40. predictive, predictor*, detect*, surveillance
41. denver developmental screening test
42. reynell, fluharty
43. development inventory
44. parent* checklist*
45. parent* survey*
46. parent* questionnaire*
47. 36,37,38,39,40,41,42,43,44,45,46
48. exp EVALUATION STUDIES
49. PREDICTIVE VALUE OF TESTS
50. "SENSITIVITY AND SPECIFICITY"
51. REPRODUCIBILITY OF RESULTS
52. exp STATISTICAL ANALYSIS
53. predictive
54. concurrent*
55. valid*
56. reliab*
57. standardi*
58. sensitivity
59. specificity
60. 48,49,50,51,52,53,54,55,56,57,58,59
61. 47 AND 21
62. 61 AND 60
63. 62 NOT 28
64. 63 NOT 35 **screening evaluation studies**
65. SPEECH THERAPY
66. SPEECH PATHOLOGY
67. SPEECH REHABILITATION
68. therap*
69. treat*
70. interven*
71. rehabilit*
72. train*
73. educat*
74. teach*
75. program*
76. facilitat*
77. imitation
78. stimulation
79. milieu
80. parent child interaction
81. 65,66,67,68,69,70,71,72,73,74,75,76,77,78, 79,80
82. RANDOMIZATION
83. RANDOMIZED CONTROLLED TRIAL
84. placebo*
85. random*
86. singl* blind*
87. doubl* blind*
88. singl* mask*
89. doubl* mask*
90. exp CLINICAL TRIALS
91. RESEARCH DESIGN
92. clinical* trial*
93. compar* report*
94. compar* stud*
95. compar* trial*
96. control*
97. prospectiv*
98. volunteer*
99. multiple baseline*
100. time series
101. quasi experimental, quasi-experimental
102. 82,83,84,85,86,87,88,89,90,91,92,93,94,95,96, 97,98,99,100,101
103. 21 NOT 29
104. 103 AND 81 AND 102 **therapy evaluation studies**
105. exp OUTCOME ASSESSMENT (HEALTH CARE)
106. PROGNOSIS
107. follow-up , follow up
108. natural* history
109. prognosis
110. outlook
111. future
112. prospective
113. 105,106,107,108,109,110,111,112
114. 103 AND 113 **natural history studies**
115. REVIEW LITERATURE
116. review* in TI
117. meta?analy*
118. systematic* review*
119. systematic* overview*

120. 115,116,117,118
121. 120 AND 47 AND 21 **reviews of screening**
122. 121 NOT 28
123. 120 AND 81 AND 103 **therapy reviews**

Deduplication then took place at the level of the reference manager software.

ERIC strategy

National Information Services Corporation
1966–September 1996.

NB. The searches for each section (ie. prevalence, screening, screening review, intervention, intervention review and natural history) were carried out individually due to a 30-line limit on search strategies. Therefore, the search strategy below will serve as a representation of those individual search strategies.

1. exp LANGUAGE IMPAIRMENTS or exp SPEECH IMPAIRMENTS or kw=(LANGUAGE ACQUISITION or LANGUAGE DEVELOPMENT or LANGUAGE IMPAIRMENTS or ARTICULATION [SPEECH] or DEVELOPMENTAL DISABILITIES or COMMUNICATION DISORDERS)
2. (speech OR language OR voice OR verbal) near3 (impair* OR delay* OR problem* OR disorder* OR retard* OR disabilit* OR difficult* OR handicap*)
3. kw=(CHILDREN or PRESCHOOL CHILDREN or TODDLERS or KINDERGARTEN CHILDREN or YOUNG CHILDREN)
4. (child* OR baby OR babies OR infant* OR toddler* OR preschool*)
5. #3 or #4
6. #1 or #2
7. #5 and #6
8. kw=(MENTAL RETARDATION or MENTAL DISORDERS or CEREBRAL PALSY or HEARING IMPAIRMENTS or DOWNS SYNDROME)
9. EPIDEMIOLOGY or INCIDENCE or PREVALENCE or RESEARCH METHODOLOGY or CROSS SECTIONAL STUDIES) or prevalen* or incidence or morbidity
10. (#7 and #9) not #8 **prevalence set**
11. kw=(SCREENING TESTS or LANGUAGE TESTS or SPEECH TESTS or DIAGNOSTIC TESTS or PHYSICAL EXAMINATIONS or HEALTH PROMOTION or REFERRAL or DIAGNOSIS or IDENTIFICATION)
12. denver developmental screening tests or reynell or fluharty or development inventory or screen* or tests or testing or tested or diagnos* or predictive or predictor* or detect* or surveillance
13. #11 or #12
14. kw=(PREDICTIVE VALIDITY or TEST VALIDITY or TEST RELIABILITY or TEST CONSTRUCTION) or predictive or concurrrent* or valid* or reliab* or standardi* or sensitivity or specificity
15. #13 and #14
16. (#7 and #15) not #8 **screening evaluation studies**
17. kw=LITERATURE REVIEWS or ti=review or meta analy*
18. (systematic* near3 (review* or overview*))
19. #17 or #18
20. #16 and #19 **reviews of screening**
21. kw=(AUTISM or MENTAL RETARDATION or MENTAL DISORDERS or CEREBRAL PALSY or HEARING IMPAIRMENTS)
22. kw=(EARLY INTERVENTION or INTERVENTION or REHABILITATION or SPEECH PATHOLOGY or SPEECH THERAPY) or imitation or stimulation or milieu or parent child interaction
23. therap* or treat* or interven* or rehabilit* or train* or educat* or teach* or program* or facilitat*
24. #22 or #23
25. (singl* or doubl* or trebl* or tripl*) near3 (blind* or mask*)
26. placebo* or random* or control* or prospectiv* or volunteer* or multiple baseline* or time series or quasi?experimental
27. kw=(CONTROL GROUPS or EXPERIMENTAL GROUPS or QUASI EXPERIMENTAL DESIGN or PROGRAM EVALUATION or EVALUATION METHODS or EVALUATION RESEARCH or COMPARATIVE ANALYSIS or PRETESTS POSTTESTS or PRETESTING
28. clinical* near3 trial*
29. (compar*) near3 (report* or study or studies or trial*))
30. #25 or #26 or #27 or #28 or #29
31. #24 and #30
32. (#7 and #31) not #21 **therapy evaluation studies**
33. #32 and #19 **therapy reviews**
34. kw=(CROSS SECTIONAL STUDIES or RESEARCH METHODOLOGY) or follow?up or (natural near3 history) or (prospective OR prognosis OR outlook OR future)
35. (#7 and #34) not #21 **natural history studies**

Deduplication then took place at the level of the reference manager software.

LLBA strategy

WINSPIRS (Silver Platter) 1973–December 1996.

DE denotes descriptor term
Lower case used for free text search

No.	Records	Request
1	4124	LANGUAGE PATHOLOGY IN DE
2	1458	SPEECH PATHOLOGY IN DE
3	696	ARTICULATION-DISORDERS in DE
4	367	VOICE-DISORDERS in DE
5	8940	CHILD-LANGUAGE in DE
6	427	DELAYED LANGUAGE IN DE
7	1063	STUTTERING in DE
8	83	mutism
9	48	echolalia
10	8303	(speech or language or voice or verbal) near5 (impair* or disorder* or disabilit* or problem* or retard* or handicap* or delay* or difficult*)
11	19758	#1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10
12	494	PRESCHOOL-CHILDREN in DE
13	1841	CHILDREN in DE
14	27293	child* or infant* or baby or babies or toddler* or preschool*
15	27293	#12 or #13 or #14
16	11776	#15 and #11
17	1207	prevalen* or morbidity or incidence
18	44	epidemiolog*
19	1238	#17 or #18
20	183	#19 and #16
21	130	SPEECH TESTING IN DE
22	131	DIAGNOSTIC TESTS IN DE
23	734	LANGUAGE TESTS IN DE
24	187	MEASURES-INSTRUMENTS in DE
25	28821	screen* or test* or diagnos* or predictive or predictor* or detect* or surveillance
26	15	development inventory
27	3	denver developmental screening test
28	32	reynell or fluharty
29	75	parent* near3 (checklist* or questionnaire* or survey*)
30	28957	#21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29
31	431	"TEST-VALIDITY-AND-RELIABILITY" IN DE
32	59	PSYCHOMETRIC-ANALYSIS in DE
33	669	RESEARCH-DESIGN-AND-INSTRUMENTATION in DE
34	11351	predictive or concurrent* or valid* or reliab* or standardi* or sensitivity or specificity
35	11955	#31 or #32 or #33 or #34
36	4930	#35 and #30
37	514	#36 and #16
38	1544	SPEECH THERAPY IN DE
39	2023	LANGUAGE THERAPY IN DE
40	377	PARENT-CHILD-INTERACTION in DE
41	70908	therap* or treat* or interven* or rehabilit* or train* or educat* or teach* or program* or facilitat*
42	4435	imitation or stimulation or milieu
43	426	parent child interaction
44	74093	#38 or #39 or #40 or #41 or #42 or #43
45	690	RESEARCH-DESIGN in DE
46	430	random* near5 (allocat* or assign* or control*)
47	2363	placebo* or random*
48	37	(singl* or doubl* or trebl* or tripl*) near5 (blind* or mask*)
49	20	clinical* near5 trial*
50	2909	compar* near5 (report* or stud* or trial*)
51	9667	control* or prospectiv* or volunteer*
52	155	multiple baseline* or time series or quasi?experimental
53	15047	#45 or #46 or #47 or #48 or #49 or #50 or #51 or #52
54	6914	#53 and #44
55	711	#54 and #16
56	31	LONGITUDINAL-STUDIES in DE
57	744	follow?up
58	72	natural near3 history
59	5569	prospective or prognosis or future or outlook
60	6371	#56 or #57 or #58 or #59
61	439	#60 and #16
62	43	DOWNS-SYNDROME in DE
63	1068	MENTAL-RETARDATION in DE
64	39	MENTAL-DISORDERS in DE
65	1	CEREBRAL-PALSY in DE
66	4057	HEARING DISORDERS IN DE
67	360	AUTISM- in DE
68	1146	#62 or #63 or #64 or #65
69	5413	#63 or #64 or #65 or #66 or #67
70	336	LITERATURE REVIEW IN DE,IP
71	3587	REVIEW* in TI
72	55	meta?analy*
73	28	systematic* near4 (review* or overview*)
74	3944	#70 or #71 or #72 or #73
75	29	#74 and #30 and #16
76	67	#74 and #44 and #16
77	173	#20 not #68

78	484	#37 not #68 not #77
79	536	#55 not #69 not #77 not #78
80	317	#61 not #69 not #77 not #78 not #79
81	18	#75 not (#68 or #77 or #78 or #79 or #80)
82	47	#76 not (#69 or #77 or #78 or #79)

Deduplication then took place at the level of the reference manager software.

MEDLINE strategy

Silver Platter 1966–December 1996.

Upper case denotes MeSH terms
Lower case used for free text search

1. "LANGUAGE-DISORDERS" / ALL SUBHEADINGS
2. explode "SPEECH-DISORDERS" / ALL SUBHEADINGS
3. "VOICE-DISORDERS" / ALL SUBHEADINGS
4. "COMMUNICATIVE-DISORDERS" / ALL SUBHEADINGS
5. "LANGUAGE-DEVELOPMENT-DISORDERS" / ALL SUBHEADINGS
6. #1 or #2 or #3 or #4 or #5
7. impair* or delay* or problem* or disorder* or retard* or disabilit* or difficult* or handicap*
8. speech or language or voice or verbal
9. #8 near5 #7
10. #6 or #9
11. "INFANT-" OR "CHILD-" OR "CHILD,-PRESCHOOL" / ALL SUBHEADINGS
12. child* or baby or babies or infant* or toddler* or preschool*
13. #10 and (#11 or #12)
14. explode "MORBIDITY" / ALL SUBHEADINGS
15. prevalen* or incidence or morbidity
16. #14 or #15
17. #16 and #13
18. explode "MASS-SCREENING" / ALL SUBHEADINGS
19. explode "POPULATION-SURVEILLANCE" / ALL SUBHEADINGS
20. "LANGUAGE-TESTS" / ALL SUBHEADINGS
21. screen* or test* or diagnos* or predictive or predictor* or detect* or surveillance
22. denver developmental screening test
23. reynell or fluharty
24. development inventory
25. parent* near3 (checklist* or survey* or questionnaire*)
26. #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25
27. explode "EVALUATION-STUDIES" / ALL SUBHEADINGS
28. "PREDICTIVE-VALUE-OF-TESTS"
29. "SENSITIVITY-AND-SPECIFICITY"
30. "FOLLOW-UP-STUDIES"
31. "REPRODUCIBILITY-OF-RESULTS"
32. "STATISTICS,-NONPARAMETRIC"
33. predictive or concurrent* or valid* or reliab* or standardi* or sensitivity or specificity
34. #27 or #28 or #29 or #30 or #31 or #32 or #33
35. #34 and #26
36. #35 and #13
37. "SPEECH-THERAPY" / ALL SUBHEADINGS
38. "SPEECH-LANGUAGE-PATHOLOGY" / ALL SUBHEADINGS
39. "SPEECH-PATHOLOGY" / ALL SUBHEADINGS
40. "LANGUAGE-THERAPY" / ALL SUBHEADINGS
41. therap* or treat* or interven* or rehabilit* or train* or educat* or teach* or program* or facilitat*
42. imitation
43. stimulation or milieu
44. parent child interaction
45. #37 or #38 or #39 or #40 or #41 or #42 or #43 or #44
46. "RANDOMIZED-CONTROLLED-TRIALS" / ALL SUBHEADINGS
47. PT = "RANDOMIZED-CONTROLLED-TRIAL"
48. "RANDOM-ALLOCATION"
49. random* near5 (allocat* or assign* or control*)
50. "DOUBLE-BLIND-METHOD"
51. "SINGLE-BLIND-METHOD"
52. PT = "CLINICAL-TRIAL"
53. explode "CLINICAL-TRIALS" / ALL SUBHEADINGS
54. clinical* near5 trial*
55. (singl* or doubl* or trebl* or tripl*) near5 (blind* or mask*)
56. "PLACEBOS" / ALL SUBHEADINGS
57. placebo* or random*
58. "RESEARCH-DESIGN" / ALL SUBHEADINGS
59. PT = "CONTROLLED-CLINICAL-TRIAL"
60. explode "EVALUATION-STUDIES" / ALL SUBHEADINGS
61. "FOLLOW-UP-STUDIES"
62. "PROSPECTIVE-STUDIES"
63. control* or prospectiv* or volunteer*
64. COMPARATIVE STUDY IN TI,AB,MESH
65. compar* near5 (report* or stud* or trial*)
66. multiple baseline* or time series or quasi?experimental
67. #46 or #47 or #48 or #49 or #50 or #51 or #52 or #53 or #54 or #55 or #56 or #57 or #58 or

#59 or #60 or #61 or #62 or #63 or #64 or #65 or #66
68. #45 and #13
69. #68 and #67
70. "DISEASE-PROGRESSION"
71. explode "OUTCOME-ASSESSMENT-(HEALTH-CARE)" / ALL SUBHEADINGS
72. "PROGNOSIS-"
73. follow?up
74. natural near3 history
75. prognosis or outlook or future or prospective
76. #70 or #71 or #72 or #73 or #74 or #75
77. #76 and #13
78. "DOWN-SYNDROME" IN MJME
79. "MENTAL-RETARDATION" IN MJME
80. "MENTAL-DISORDERS" IN MJME
81. "CEREBRAL-PALSY" IN MJME
82. "HEARING-DISORDERS" IN MJME
83. "AUTISM-" IN MJME
84. #78 or #79 or #80 or #81
85. #79 or #80 or #81 or #82 or #83
86. PT=REVIEW
87. "META-ANALYSIS"
88. "REVIEW-LITERATURE"
89. meta?analy*
90. #86 or #87 or #88 or #89
91. #91 and #13 and #26
92. #91 and #13 and #45
93. #17 not #84 **prevalence set**
94. #36 not #84 not #94 **screening evaluation studies**
95. #92 not #84
96. #93 or #94
97. #95 not #96 **reviews of screening**
98. #92 not #85
99. #96 or #97
100. #98 not #99 **therapy reviews**
101. #69 not #85
102. #101 not #96 **therapy evaluation studies**
103. #77 not #85
104. #101 or #96
105. #103 not #104 **natural history studies**

Deduplication then took place at the level of the reference manager software.

PsychLit strategy

Silver Platter: 1974–December 1996 (journals); 1987–December 1996 (chapters and books)

DE denotes descriptor term
Lower case used for free text search

1. exp LANGUAGE DISORDERS in DE
2. LANGUAGE DEVELOPMENT in DE
3. LANGUAGE DELAY in DE
4. exp SPEECH DISORDERS in DE
5. SPEECH DEVELOPMENT in DE
6. SPEECH HANDICAPPED in DE
7. EARLY CHILDHOOD DEVELOPMENT in DE
8. DELAYED DEVELOPMENT in DE
9. (speech OR language OR voice OR verbal) NEAR5 (impair* OR disorder* OR disabilit* OR problem* OR retard* OR handicap* OR delay* OR difficult*)
10. (1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9)
11. 10 AND AG=CHILD
12. 10 AND (child* OR infant* OR preschool* OR baby OR babies OR toddler*)
13. 11 OR 12
14. DOWN SYNDROME in DE
15. CEREBRAL PALSY in DE
16. MENTAL RETARDATION in DE
17. MENTAL DISORDERS in DE
18. HEARING DISORDERS in DE
19. AUTISM in DE
20. 14 OR 15 OR 16 OR 17
21. 15 OR 16 OR 17 OR 18 OR 19
22. EPIDEMIOLOGY in DE
23. exp METHODOLOGY
24. (prevalenc* OR morbidity OR incidence)
25. 22 OR 23 OR 24
26. 13 NOT 20
27. 26 AND 25 **prevalence set**
28. SCREENING in DE
29. SCREENING TESTS in DE
30. PHYSICAL EXAMINATION in DE
31. HEALTH EDUCATION in DE
32. HEALTH PROMOTION in DE
33. PROFESSIONAL REFERRAL in DE
34. PROFESSIONAL CONSULTATION in DE
35. DIAGNOSIS in DE
36. AT RISK POPULATIONS in DE
37. (screen* OR test* OR diagnos* OR predictive OR predictor* OR detect* OR surveillance)
38. denver developmental screening test
39. reynell OR fluharty
40. development inventory
41. parent* NEAR3 (checklist* OR survey* OR questionnaire*)
42. 28 OR 29 OR 30 OR 31 OR 32 OR 33 OR 34 OR 35 OR 36 OR 37 OR 38 OR 39 OR 40 OR 41
43. PREDICTIVE VALIDITY in DE
44. STATISTICAL VALIDITY in DE
45. TEST VALIDITY in DE
46. TEST RELIABILITY in DE
47. TEST CONSTRUCTION in DE
48. TEST STANDARDIZATION in DE
49. (predictive OR concurrent* OR valid* OR reliab* OR standardi* OR sensitivity OR specificity)

50. 43 OR 44 OR 45 OR 46 OR 47 OR 48 OR 49
51. 42 AND 50
52. 51 AND 26 **screening evaluation studies**
53. LITERATURE-REVIEW in DE
54. META-ANALYSIS in DE
55. review* in TI
56. meta?analy*
57. (systematic* NEAR4 (review* OR overview*))
58. 53 OR 54 OR 55 OR 56 OR 57
59. 42 AND 58
60. 59 NOT 52 **reviews of screening**
61. EARLY INTERVENTION in DE
62. INTERVENTION in DE
63. SPEECH THERAPY in DE
64. TREATMENT in DE
65. REHABILITATION in DE
66. PARENT TRAINING in DE
67. PARENT CHILD INTERACTION in DE
68. (therap* OR treat* OR interven* OR rehabilit* OR train* OR educat* OR teach* OR program* OR facilit*).tw
69. (imitation OR stimulation OR milieu).tw
70. parent child interaction.tw
71. 61 OR 62 OR 63 OR 64 OR 65 OR 66 OR 67 OR 68 OR 69 OR 70
72. (singl* OR doubl* OR trebl* OR tripl*) NEAR5 (blind* OR mask*)
73. (clinical* NEAR5 trial*)
74. PLACEBO in DE
75. (placebo* OR random*)
76. exp EXPERIMENTAL DESIGN in DE
77. EVALUATION in DE
78. PROGRAM EVALUATION in DE
79. TREATMENT EFFECTIVENESS EVALUATION in DE
80. EXPERIMENT CONTROLS in DE
81. PRETESTING in DE
82. (compar* NEAR5 (report* OR stud* OR trial*))
83. (control* OR prospectiv* OR volunteer*)
84. (multiple baseline* OR time series OR quasi?experimental)
85. TIME SERIES in DE
86. 72 OR 73 OR 74 OR 75 OR 76 OR 77 OR 78 OR 79 OR 80 OR 81 OR 82 OR 83 OR 84 OR 85
87. 71 AND 86
88. 13 NOT 21
89. 88 AND 87
90. 89 NOT 52 **therapy evaluation studies**
91. 88 AND 71 AND 58
92. 91 NOT (52 OR 90 OR 60) **therapy reviews**
93. follow?up
94. natural NEAR3 history
95. (prospective OR prognosis OR outlook OR future)
96. FOLLOW-UP STUDIES in DE
97. 93 OR 94 OR 95 OR 96
98. 97 AND 88 **natural history studies**

Deduplication then took place at the level of the reference manager software.

Calls for information

Overseas associations of speech and language therapists

Organisations contacted in order to publish announcement of project in relevant organisational bulletin, and to ask for register of current research or list of known reviews:

American Speech-Language Hearing Association*
Asociacion Espanola de Logopedia, Foniatria y Audiologia (AELFA)*
Canadian Association of Speech-Language Pathologists and Audiologists
Dutch Society of Logopedics and Phoniatrics (Nederlandse Verening Voor Logopedie en Foniatria, NVLF)*
German Association of Logopedics*
Speech Pathology Association of Australia*
New Zealand Speech-Language Therapists Association*
The Hong Kong Association of Speech Therapists

* Individuals contacted the review after seeing the project announced.

UK institutions training speech and language therapists

The following 15 teaching institutions were sent an information sheet about the review. Work in progress or unpublished work specifically requested.

Central School of Speech and Drama, London
College of St. Mark and St. John, Devon
De Montfort University, Leicester
Department of Human Communication Science, University College, London*
Leeds Metropolitan University
Manchester Metropolitan University
Queen Margaret College, Edinburgh*
University of Central England in Birmingham
University of Manchester
University of Newcastle-upon-Tyne
University of Reading
University of Sheffield
University of Strathclyde, Glasgow*

University of Ulster at Jordanstown,
County Antrim, Ireland
University of Wales Institute Cardiff

* Individuals from these institutions made submissions to the review.

Royal College of Speech and Language Therapists

Announcement in Bulletin of RCSLT (September 1996) as call for papers. (The RSCLST did not at the time of the review hold an up-to-date register of research in progress in UK.)

Agencies giving advice

Outcomes Clearing House, Nuffield Institute for Health, Leeds
Centre for Evidence-based Medicine, Oxford
Special Needs Research Unit, University of Northumbria
National Centre for Clinical Audit, British Medical Association, London

Health Visitor Association

Request to HV magazine to announce project.

E-mail bulletin boards

(Estimated number of delivery destinations in brackets where known.)

Psychology departments at UK universities: AHPD@psyc.leeds.ac.uk (42)
Education departments at UK universities: bera@vax.sbu.ac.uk
Language development researchers (CHILDES group; Child Language Data Exchange): info-childes@andrew.cmu.edu (211)
Members of the Centre for Evidenced-based Medicine Group: cebm-members@mailbase.ac.uk

Current researchers

As identified through the National Research Register (August 1996, search courtesy of the CRD).

Ms H Alton, NW Anglia Health Authority
Ms V Joffe, University of Oxford
Dr S Roulstone, Frenchay Hospital, Bristol*
Ms L Smith, Kent and Canterbury Hospital*
Ms C Stott, CLASP project, University of Cambridge*
Ms M Trim, Aylesbury Vale Healthcare NHS Trust*
Ms N Warrick, University of Oxford*

* These researchers responded with copies of work or a discussion of their progress.

Conference proceedings

CPLOL, Lisbon, Portugal; May 3–5, 1997.
Abstracts for this conference were scanned, and selected authors contacted for more information:

E Roy, Nancy, France*
D Tavares, Centro de Medicina de Reabilitacao do Alcoitao, Portugal*
M Vlassopuolos, Athens, Greece
M Westerlund, Uppsala University Children's Hospital, Sweden*

* These researchers responded with references and reprints.

Authors of articles relevant to screening or therapy evaluation

These selected on basis of initial handsearch of journals, or as being leading names in the field.

Adesman A, Schneider Children's Hospital, USA
Anderson C, Speech and Language Therapist, UK*
Attanasio JS, Montclair State College , USA*
Aylward G, SIU School of Medicine, USA
Badenoch D, NHS R&D Centre for Evidenced-Based Medicine, UK*
Bates E, USA
Bennett FC, University of Washington, USA
Berger S, Evanston Hospital, USA
Best W, Birkbeck College, UK*
Blackman J, Kluge Children's Rehabilitation Center, USA
Boyce WT, University of California, USA
Bozic N, University of Birmingham, UK*
Braden R, University of Colorado, USA
Camarata S, Vanderbilt University, USA
Camp B, University of Colorado, USA
Capute A, The Johns Hopkins University School of Medicine, USA*
Carey W, Children's Hospital, Philadelphia, USA
Chesky R, USA
Cole KN, University of Washington, USA
Cooper M, Plymouth Community Services Health Trust, UK*
Coplan J, SUNY Health Science Center, USA
Cordes AK, University of California, USA
Coury D, Children's Hospital, Columbus, USA
Dale PS, University of Washington, USA
Damico J, University of Southwestern LA, USA
Dancer J, Speech and Language Therapist, UK
Deeley W, Doncaster Healthcare NHS Trust, UK
Dixon S, UC Medical Center, USA
Dodd B, University of Newcastle-upon-Tyne, UK
Donaldson M, University of Edinburgh, UK*
Durlak J, Loyola University Chicago, USA*
Elbert M, Indiana University, USA
Feldman H, Children's Hospital of Pittsburgh, USA*

Felsenfield S, University of Pittsburgh, USA
Felt B, Center for Human Growth and Development, Michigan, USA*
Fey ME, University of Kansas Medical Center, USA*
Field N, Chorley and South Ribble Health Authority, UK
Filler MD, James Madison University, USA
Fox MA, Thames Valley Children's Centre, USA
Friedman S, Montefiore Medical Center, USA
Friel-Patti S, UTD/Callier Center for Communication Disorders, USA
Garcia-Tommel S, Hospital Sant Joan de Deu, Spain
Gibbard D, SALT Dept, Havant Health Centre, UK
Gillberg C, Annedals clinics, Goteburg, Sweden
Grieve R, Edinburgh University, UK
Hewitt A, North Mersey Community Health Trust, UK*
Hodell S, Epsom Health Care Trust, UK*
Hoffman PR, Louisiana State University, USA
Howlin P, St. George's Hospital, London, UK*
Hyde Wright S, Speech and Language Therapist, Dawn House School, UK*
Iacono T, University of Nebraska-Lincoln, USA
Ingham JC, University of California, USA
Ireton H, University of Minnesota, USA*
Jacklin A, University of Sussex, UK
Jellinek M, USA
Johnston JR, University of British Columbia, Canada
Kaiser AP, Vanderbilt University, USA*
Kelleher K, University of Pittsburgh, USA
Kellow B, Speech and Language Therapist, UK*
Kuhn T, Mecklenburg Center for Human Development, USA
Kwaitkowski J, University of Wisconsin-Madison, Madison, USA
Lahey M, USA
Law P, University of Vermont, USA
Levine MD, University of North Carolina, USA
Lewis V, School of Education, Open University, UK*
Lord C, Department of Psychiatry, Chicago, USA
Magill S, North Downs and Ards Community HSS Trust, UK*
McDade A, Speech and Language Therapist, UK*
McEvoy RE, University of Colorado, Health Sciences Center, USA
McGregor K, Northwestern University, USA
McGurk H, Australian Institute of Family Studies, Australia
Melhuish EC, Department of Psychology, University College of North Wales, UK
Menyuk P, Boston University, USA
Miller JF, University of Wisconsin-Madison, USA
Muma, J, Central Michigan University, USA*
Murray L, University of Cambridge, UK
Murray-Branch J, University of Wisconsin-Madison, USA
Nightingale S, ICAN, UK
Norris J, Louisiana State University, USA*
Oberklaid F, Centre for Community Child Health, Victoria, Australia
Onslow M, University of Sydney, Australia
Palfrey JS, Children's Hospital, Boston, USA
Pantell R, University of California, USA*
Parker N, Merrylands CHC, New South Wales, Australia
Parker S, Boston City Hospital, USA
Paul R, Department of Speech Communication, Portland State University, USA
Rescorla L, Bryn Mawr College, USA
Ruben JJ, Albert Einstein College of Medicine, New York, USA*
Scarborough H, Brooklyn College, New York, USA
Schechter N, St Francis Hospital, Hartford, USA
Schuler A, San Francisco State University, USA
Schwartz HD, North Illinois University, USA
Scott J, Burnley Health Trust, UK
Secord W, North Arizona University, USA
Shriberg LD, University of Wisconsin-Madison, USA
Silver K, Institute of Education, UK
Squires J, University of Oregon, USA*
Stancin T, Metro Health Medical Center, USA
Stein M, UCSD School of Medicine, USA*
Stern LM, Regency Park Centre for Young Disabled, Australia*
Stevenson J, Institute of Child Health, University of London, UK
Sturner RA, The Johns Hopkins University School of Medicine, USA*
Tager-Flusberg H, University of Massachusetts, USA
Thal DJ, San Diego State University, USA
Tomblin JB, University of Iowa, USA*
Trimm F, University of South Alabama, USA
von Benda U, Max Planck Institute for Psychiatry, Germany
Warren SF, Vanderbilt University, USA*
Warrick N, University of Oxford, UK*
Wasserman R, University of Vermont, USA
Watson MM, Central Michigan University, USA
Watt J, Edinburgh University, UK
Weismer SE, University of Wisconsin-Madison, USA
Wender E, Child Health Services, USA
Whitehurst GJ, State University of New York, USA*
Wiig E, USA
Williamson I,Aldermoor Health Centre, Southampton, UK
Wing CS, USA
Wong V, University of Hong Kong, Hong Kong*
Yairi E, University of Illinois, USA
Yoder PJ, Vanderbilt University, USA

* These researchers responded with names, references or reprints.

Individuals contacting the review team

Relevant information was submitted by the following people in response to hearing about the review.

Dejonckere PH, Academisch Ziekenhuis Utrecht, The Netherlands
Fisher J, Royal Children's Hospital, Melbourne, Australia
Gibbon F, Speech and Language Therapist, UK
Glascoe FP, Vanderbilt University, USA
Goorhuis-Brouwer S, Academisch Ziekenhuis Groningen, The Netherlands
Klee T, University of Newcastle-upon-Tyne, UK
Pozsonyi V, Speech and Language Therapist, UK
Ridder-Sluiter H, Nederlandse Stichting voor het Dove en Slechthorende Kind, The Netherlands

Appendix 4

Inclusion and exclusion criteria

Papers meeting all of the criteria for their study type in terms of both relevance and construct, internal, external or statistical conclusion validity were included in the data extraction stage.

Relevant papers failing on aspects of validity were logged as 'excluded papers', without being included in the data extraction stage.

Prevalence studies

Criteria	Include	Exclude
1. Reported after 1966 *(relevance)*	Yes	No
2. Study of the prevalence of speech and language delay in a sample that includes children aged up to 16 years *(relevance)*	Yes	No
3. Study is of primary speech and language delay *(relevance):* Is the focus of the study confined to one or more of the following (circle): ADD/ADHD; deafness/sensorineural loss; autism; psychiatric or EBD; Down's Syndrome; cerebral palsy; dyslexia; other secondary speech and language delay (specify: e.g. mental retardation, mental handicap, neurological conditions such as epilepsy, closed head injury – specify others)	No	Yes
4. Information about the sample size seen for full diagnostic testing *(statistical conclusion validity, external validity)*	Yes	No
5. Is the sample drawn from a general population (i.e. excluding 'clinic' samples or numbers referred to hospitals)? *(external validity)*	Yes	No
6. Do the clinicians or researchers use clear criteria for defining speech and language delay? *(internal validity)*	Yes	No

Natural history/follow-up studies of non-systematic intervention

Criteria	Include	Exclude
1. Reported after 1966 *(relevance)*	Yes	No
2. Longitudinal study of a population/sample of children identified as having speech and language delay who have NOT received any specific treatment or intervention for the target condition (natural history) or who have received only non-systematic treatment or intervention (follow-up) (e.g. general, non-systematic advice only) *(acceptability)*	Yes	No
3. Children identified as having speech and language delay before the age of 7 years *(relevance)*	Yes	No
4. Study is of primary speech and language delay *(relevance):* Is the focus of the study confined to one or more of the following (circle): ADD/ADHD; deafness/sensorineural loss; autism; psychiatric or EBD; Down's Syndrome; cerebral palsy; dyslexia; effects of social disadvantage; other secondary speech and language delay (specify: e.g. mental retardation, mental handicap, neurological conditions such as epilepsy, closed head injury – specify others)	No	Yes
5. Prospective study *(statistical conclusion validity, external validity)*	Yes	No
6. Specified follow-up interval of 6 months or more *(internal validity, statistical conclusion validity)*	Yes	No
7. Pre-test and post-test language measures available (norm-referenced or criterion-referenced) *(internal validity)*	Yes	No

Intervention studies

Criteria	Include	Exclude
1. Reported after 1966 *(relevance)*	Yes	No
2. Covers part of the age-range 0–7 years *(relevance)*	Yes	No
3. Study of the effects of treatment/intervention upon speech and language delay in children *(relevance)*	Yes	No
4. Study is of primary speech and language delay *(relevance):* Is the focus of the study confined to one or more of the following: ADD/ADHD; deafness/sensorineural loss; autism; psychiatric or EBD; Down's Syndrome; cerebral palsy; dyslexia; effects of social disadvantage; other secondary speech and language delay (specify: e.g. mental retardation, mental handicap, neurological conditions such as epilepsy, closed head injury – specify others)	No	Yes
5. Details of number of participants in each group *(statistical conclusion validity, external validity)*	Yes	No
6. Provides a comparison of pre- and post-intervention speech and language measures (e.g. mean, SDs, and/or details of any statistical procedures including percentages) *(internal validity, statistical conclusion validity)*	Yes	No
7. Fulfils one of the following design criteria chosen to minimise threats to internal validity (indicate which) *(internal validity):* (a) experimental study with randomised non-treatment controls; (b) quasi-experimental studies (with non-random/pseudo-random or non-equivalent non-treatment control groups) including interrupted multiple time series study (with non-equivalent non-treatment controls); (c) the following single-subject experimental designs: withdrawal and reversal designs (ABAB, BAB, ABA), multiple baseline designs (across behaviours, settings or subjects), multiple probe designs or alternating treatment designs, all with graphical displays or session-by-session data for individuals (baselines should have > 2 points with the exception of multiple baseline designs where one of the baselines may have 2 points)	Yes	No
8. Details of nature, duration, span and delivery of treatment *(construct validity; internal validity)*	Yes	No
If excluded indicate direction of outcomes for treatment/intervention: (circle)		+ve null –ve

Screening tests and procedures or associated validation studies

Criteria	Include	Exclude
1. Reported after 1966 *(relevance)*	Yes	No
2. Tests contain speech and language items *(relevance)*	Yes	No
3. Test designed to be used in (a) a primary health care setting and/or (b) in an educational setting by non-specialist staff for early identification, not diagnosis (relevance)	Yes	No
4. Covers part of the age-range 0–7 years *(relevance)*	Yes	No
5. Study is of primary speech and language delay *(relevance):* Is the focus of the study confined to one or more of the following (circle): ADD/ADHD; deafness/sensorineural loss; autism; psychiatric or EBD; Down's Syndrome; cerebral palsy; dyslexia; other secondary speech and language delay (specify: e.g. mental retardation, mental handicap, neurological conditions such as epilepsy, closed head injury – specify others)	No	Yes
6. Information about the sample size *(statistical conclusion validity, external validity)*	Yes	No
7. Is the sample drawn from either (a) a general population or (b) a clinical population of children with speech and language problems, or (c) a mixed clinical/general population? *(external validity)*	Yes	No
8. Are there clear criteria for speech and language delay based on cut-off scores on gold-standard norm-referenced tests or objectified clinical judgement? *(internal validity)*	Yes	No
9. Is information given which allows calculation of concurrent validity for the speech and language items? *(statistical conclusion validity)*	Yes	No
10. Is concurrent validity expressed by both sensitivity and specificity? *(statistical conclusion validity)*	Yes	No

Appendix 5

Summary of reviewed studies

TABLE 27 *Prevalence studies: pre-school (up to 5 years)*

Study	Criteria for language delay	Population sampled		Prevalence		Ranking			Comments
	Reference test, cut-off	**No. given full testing/ screened sample**	**Age (years; months)**	**Prevalence**	**CI**	**Replic-ability (of 9)**	**Validity (of 12)**	**Total (of 21)**	
Bax *et al,* 1983 with Bax *et al,* 1980 Camden and Westminster, London, UK	Clinical judgement supported by a reference test: a. 'abnormal' speech and/or language b. 'possibly abnormal' speech and/or language.	296/304 323/333 269/278	2;0 3;0 4;6	a. 5% b. 17% a. 8% b. 12% a. 5% b. 7%	Not stated	5	6	11	Children attending routine developmental checks; catch-ment areas of three clinics. Negative impact of low SES noted. Reference test supports use of clinical judgement: verified on approximately 60 children.
Burden *et al,* 1996 Cambridge, UK	Renfrew Action Picture Test (RAPT); Renfrew Bus Story; devised test of receptive language; devised test of speech. Cut-offs: one score below 5th centile, **or** at least three scores at 5–8th centiles.	425/1936	3;3	Speech and/ or language delay 6.9% Males 10.6% Females 4.6%	± 2.86	6	11	17	Use of own norms for setting centiles on the Renfrew and devised tests; 10th centile on their sample estimated as 5th centile for general population, which is format quoted here. Prevalence may have been underestimated due to non-responders.
Randall *et al,* 1974 Barnet, London, UK	Reynell Developmental Language Scales (RDLS); Articulation task (picture naming); intelligibility of speech rated 1–4. Cut-off: –2 SD on any of the tests.	176/176	2;11–3;02	Severely speech retarded 5.6% Males 6.25% Females 5%	Not stated	7	12	19	Prevalence based on 160 English-speaking families of the 176.
Rescorla, 1993 Philadelphia, USA	1. Bayley objects (ball, cup, clock, pencil, scissors), none named. 2. 14 pictures from the Stanford Binet Intelligence Scale, vocabulary sub-test; none named.	92/92	Mean 2 years	Expressive language delay Delay 1: 8% Less than 30 words **and** no word combinations Delay 2: 16% Less than 30 words **or** no word combinations Delay 3: 19% Less than 50 words **or** no word combinations	Not stated	7	8	15	Rescorla quotes the prevalence of expressive language delay according to three cut-offs on the Language Development Survey screen. The paper implies prevalence rates of 9.8% (according to the Bayley cut-off) and 13% (according to the Binet cut-off). Sample high SES: only 15% from classes III, IV, V.

continued

TABLE 27 contd *Prevalence studies: pre-school (up to 5 years)*

Study	Criteria for language delay	Population sampled		Prevalence		Ranking			Comments
	Reference test, cut-off	No. given full testing/ screened sample	Age (years; months)	Prevalence	CI	Replic-ability (of 9)	Validity (of 12)	Total (of 21)	
Silva *et al*, 1983 with Silva *et al*, 1980 Dunedin, New Zealand	RDLS; receptive and expressive: a. expressive delay – less than 5th centile on only an expressive scale b. receptive delay – less than 5th centile on only a receptive scale c. expressive/receptive (general language delay) – less than 5th centile on both an expressive and receptive scale.	1027/1027	3;0	Any language delay 7.6% a. expressive delay 2.34% b. receptive delay 2.63% c. general language delay 2.63%	Not stated	6	12	18	Relatively high SES. See later figures for 5 and 7 years.
Stewart *et al*, 1986 Washington, USA	Goldman-Fristoe Test of Articulation (GFTA), cut-off not stated; Black English productions allowed; Peabody Picture Vocabulary Test (PPVT), vocabulary more than 1 year below CA; Utah Test of Language Development (TOLD), scored in usual manner; language sample inspected for maturity with respect to developmental norms.	65/719	3–5 years	speech 1.5% language 2.6%	Not stated	7	4	11	Only gave follow-up testing to some pre-screen fails, thus underestimating prevalence. Sample 100% black.
Stevenson and Richman, 1976 Walthamstow, London, UK	RDLS (expressive): a. expressive language delay – cut-off, LA less than 30 months (i.e. 6-month delay) b. severe expressive language delay – cut-off, LA less than two-thirds of CA c. specific expressive language delay not associated with general retardation – cut-off, LA less than two-thirds of mental age, and mental age above two-thirds of CA.	205/705	3 years	a. 3.12% Males 4.38% Females 1.93% b. 2.27% Males 3.22% Females 1.38% c. 0.57% Males 0.88% Females 0.27%	1.84–4.4 1.17–3.37 0.02–1.12	6	10	16	Checked the adequacy of his screening procedure in order to find as many true cases as possible. Excluded non-indigenous families (resident in UK less than 20 years).
Wong *et al*, 1992 Hong Kong	RDLS (Cantonese version), expressive and receptive scales: a. severe language delay – cut-off, LA less than two-thirds of CA b. specific language delay not associated with general retardation – cut-off, LA less than two-thirds of mental age, and mental age above two-thirds of CA.	233/855	3 years	a. 3.4% Males 2.7% Females 0.9% b. 3% Males 2.2% Females 0.9%	Not stated	6	12	18	Definition of delay and cut-offs follow Stevenson and Richman, except that Wong *et al* use combined expressive and receptive scales of RDLS. Prevalence of 3.4% is based on a diagnostic sample of 233; elsewhere in the paper a sample of 226 is cited.

continued

TABLE 28 *Prevalence studies: school age (5 years and above)*

Study	Criteria for language delay	Population sampled		Prevalence		Ranking			Comments
	Reference test, cut-off	No. given full testing/ screened sample	Age (years; months)	Prevalence	CI	Replic-ability (of 9)	Validity (of 12)	Total (of 21)	
Beitchman *et al*, 1986 Ottawa Carleton Region, Canada	TOLD; –1 SD (SLQ sub-test) –2 SD (any test) PPVT-R; –1 SD Goldman-Fristoe-Woodcock auditory memory tests; below 15th centile on both tests. Checklist for voice, stuttering, dysarthria problems; cut-off not given.	352/1655	5 years	Speech only 6.4% Males 6.58% Females 6.68% Language only 8.04% Males 8.17% Females 8.37% Speech and language 4.56% Males 3.31% Females 7.06% Speech or language 19.0% Males 18.1% Females 22.1%	Not stated Not stated Not stated ± 2.8%	7	12	19	High number of cases found within screen passes results in observed prevalence of 11% moving to a corrected prevalence of 19.0%. See text for comment on male:female ratio Using a –2 SD cut-off * on the PPVT and the TOLD SLQ: speech only: 8.6 language only: 4.3 speech and language: 3.1 speech or language: 12.1 * (Personnal communication)
Dudley and Delage, 1980 Quebec, Canada	8 standardised measures of receptive and expressive language; (7 French tests and PPVT; 1 measure of verbal memory; –1.65 SD on any one test.	334/334	Nursery, Year 1 and Year 2	Speech (L'articulation) N: 7.8% Y1: 12.6% Y2: 2.3% Language (langage) N: 9.1% Y1: 5.5% Y2: 3.1%	Not stated	4	7	11	Tests given in random order. Clinical judgement of two others used to 'evaluate' the test scores.
Harasty and Reed, 1994 Sydney, Australia	Fisher-Logemann Test of Articulation Competence, guidelines given. TOLD (Primary/ Intermediate/Adolescent); –1 SD on SLQ or –2 SD on any sub-test. Pragmatics protocol (Prutting and Kirchner), 30% level of inappropriacy. Voice and stuttering checklists, guidleines given.	70/437	Kinder-garten to Grade 6	Speech (including articulation, voice and fluency): 12.6% Language only: 12.6% Combined speech and language impairment: 8.0% Overall estimate of any impairment: 33.2%	± 3.1% ± 3.1% ± 1.3% ± 4.4%	5	7	12	High attrition (due to refusals) when making up the diagnostic sample. Higher prevalence noted for students of a non English-speaking background.
Kirkpatrick and Ward, 1984 New South Wales, Australia	Edinburgh Articulation Test (EAT), –2 SD from mean error score at their grade level.	2251/2251	Kinder-garten to Grade 1	Speech 4.6%	Not stated	6	9	15	

continued

TABLE 28 contd *Prevalence studies: school age (5 years and above)*

Study	Criteria for language delay	Population sampled		Prevalence		Ranking			Comments
	Reference test, cut-off	**No. given full testing/ screened sample**	**Age (years; months)**	**Prevalence**	**CI**	**Replic-ability (of 9)**	**Validity (of 12)**	**Total (of 21)**	
Paul *et al,* 1992 with Thorburn 1991 Clarendon, Jamaica	Clinical judgement based on outlined categories of expressive language and speech: mild, moderate and severe.	995/5458	2–9 years	All levels of severity 1.35% Moderate/ severe ('serious'): 0.65% Severe: 0.31%	Not stated	3	12	15	Sampling and subsampling procedures are strong; however, speech/language diagnostic crtierion does not readily lend itself to replication.
Silva *et al,* 1983 with Silva *et al,* 1980 Dundedin, New Zealand	RDLS; receptive and expressive: a. expressive delay – less than 5th centile on only an expressive scale b. receptive delay – less than 5th centile on only a receptive scale c. expressive/receptive (general language delay): less than 5th centile on both an expressive and receptive scale.	936/1027	5 years	Any language delay 10.4% a. expressive only delay 4.27% b. receptive only delay 3.95% c. expressive/ receptive language delay 2.14%	Not stated	6	12	18	Relatively high SES.
	Illinois Test of Psycholinguistic Abilities (ITPA) for auditory reception and verbal expression, less than 5th centile.	891/1027	7 years	Any language delay 8.4% a. expressive only delay 2.81% b. receptive only delay 3.59% c. expressive/ receptive language delay 2.02%					
Tomblin, unpublished (with Tomblin *et al,* 1996) Iowa, USA	Seven standardised language measures, giving five composite language scores; at least two of five composite scores –1.25 SD below age group norms.	2009/7218	5 years	Specific language impairment 7.4%	6.8–8.5%	6	11	17	In addition to language criteria, specific language impaired children had to satisfy: – normal unilateral hearing at 500 Hz, 1 kHz and 2 kHz – normal non-verbal performance. Tomblin discusses issue of cut-offs affecting prevalence estimates.
Warr-Leeper *et al,* 1979 Oklahoma, USA	Hejna Developmental Articulation Test; reading passage; voice profile; cut-offs not stated.	170/999	Grades 6 to 8	Articulation problems or articulation and vocal problems 7.3%	Not stated	1	6	7	Screening measures repeated as a diagnostic if result suspect. No checking of false-negatives.

continued

TABLE 28 contd *Prevalence studies: school age (5 years and above)*

Study	Criteria for language delay	Population sampled		Prevalence		Ranking			Comments
	Reference test, cut-off	**No. given full testing/ screened sample**	**Age (years; months)**	**Prevalence**	**CI**	**Replic-ability (of 9)**	**Validity (of 12)**	**Total (of 21)**	
Tuomi and Ivanoff, 1977 London; Ontario, Canada	Speech. Tests from Fisher-Logemann Test of Articulation; Templin-Darley tests; GFTA; MacDonald Deep Test of Articulation. Poor stimulability and multiple, consistent errors. Language. Tests from Northwestern Syntax Screening Test (NSST); PPVT; Test of Auditory Comprehension of Language (TACL); ITPA grammatical closure sub-test. More than one year behind for age.	899 screened; screen fails given diag-nostic testing, number unspecified	Kinder-garten to Grade 1	Speech (excluding voice, fluency): Kinder-garten 24.6% Grade 1 16.5% Language: Kinder-garten 6.2% Grade 1 7.1%	Not stated	2	6	8	Gender balance ratios given: 2.3:1 for male:female for speech (Kindergarten and Grade 1); 1.6:1 for language (Kindergarten) and 1.2:1 (Grade 1). Not clear if designation for therapy also signifies definition of a true case. Only screen fails seen for further testing; no checking of false-negatives.

TABLE 29 *Natural history studies*

Study	Population sampled	Follow-up duration and measures	Study conclusions	Ranking			Comments
	Subjects; recruitment; criteria for language delay	**Duration; measures; age at follow-up**	**Numbers of children remaining delayed; other observations**	**Replicability (of 12)**	**Validity (of 19)**	**Total (of 31)**	
Bralley and Stoudt, 1977 USA	n = 60 Initial age approximately 6–7 years: grade 1 entry. Pupils in selected schools known to be without speech and language therapy services. Criteria for speech problem: at least one error on the Arizona Articulation Proficiency Scale (AAPS).	Followed-up annually for 5 years. AAPS repeated each year. Age at end of study approximately 11–12 years.	1. 13 of 60 (21.6%) children persisted in their speech problems to beginning of grade 5. Of these two were severe cases. 2. 30% of cases had persisted to beginning of grade 4. 3. Grade 1 mean number of errors per child: 6.6; Grade 5: mean number of errors per child: 1.3. 4. Agreement noted with authors van Riper and Erickson (1968) who stated most children spontaneously recover articulation errors by end grade 3.	5	17	22	No breakdown of male:female ratios. Speech data only, not language. Criterion-referenced measure.
Felsenfeld *et al*, 1992 USA	n = 24 Initial age 56–60 months (kindergarten age) Control group of 28 subjects, not matched at outset but selected at follow-up. Recruitment: selected subjects from 1960–1972 Templin study of articulation. Criteria for speech problem: a. at least 1.5 SD below sample mean on first five assessment results of Prekindergarten Imitation Articulation Test (Templin, 1960) b. no organic impairment c. no hearing loss d. no mental retardation.	Follow-up age 32–34 years (approximately 28-year follow-up). Measures at follow-up: a. articulation: Templin-Darley Test of Articulation and Screening Test; fail if raw score less than 76. b. expressive language: vocabulary and comprehension sub-tests of Wechsler Adult Intelligence Scale – Revised c. receptive language: PPVT and revised Token Test; fail if below 1 SD d. cognitive skills: block peformance sub-tests of Wechsler Adult Intelligence Scale – Revised e. personality scale: Eysenck Personality Inventory.	1. Speech: 50% failed speech task. 2. Receptive language: early speech-delayed group performing at 1.2 SD below mean of controls; also 33% failed the Token Test. 3. Expressive language: early speech-delayed group performing at 0.82 and 0.89 SD below mean of controls on tasks of verbal reasoning and vocabulary. 4. Non-verbal reasoning scores within normal range; personality scores not significantly different from controls. Thus children primarily speech delayed at preschool experience persistent problems.	7	10	17	Follow-up sample a subset of Templin's 1960 longitudinal study; original population sample of 1500. Subjects not eligible for treatment until grade 3 (approximately 8 years). No specific reference to any later therapy received by the subjects. Mixture of norm-referenced and non norm-referenced measures.

continued

TABLE 29 contd *Natural history studies*

Study	Population sampled	Follow-up duration and measures	Study conclusions	Ranking			Comments
	Subjects; recruitment; criteria for language delay	**Duration; measures; age at follow-up**	**Numbers of children remaining delayed; other observations**	**Replicability (of 12)**	**Validity (of 19)**	**Total (of 31)**	
Fiedler *et al*, 1971 USA	n = 46 (language delayed), with 92 controls (matched for race, gender, SES). Initial age 3 years. Recruitment: all mothers and their children attending a Boston Hospital for antenatal care. Criteria for language delay: failure at 3 years on a speech/language screening task.	1 and 4 year follow-up, at ages 4 and 7 years. Measures: a. psychological and neurological measures b. examiner comment on any speech abnormality.	1. At age 4, 87% persisted, while at 7 years 38% persisted in showing poor speech/language development. 2. Of the control group, 29% and 8% emerged as having minor speech problems at ages 4 and 7 years. 3. For the 3-year language-delayed group, higher rates of abnormality observed at 7 years on the psychological and neurological measures than for controls.	4	10	14	Pre- and perinatal data also recorded with infant developmental details. Risk factors for poor speech/language identified in perinatal factors and first year development. Very limited information on speech/language assessment measures used. Nature of speech/language deficits not described. Criterion-referenced measures.
Hall, 1996 with Hall *et al*, 1993 USA	n = 5 (untreated) Initial age 41–70 months. Recruitment: selected from earlier study sample (high dysfluency subgroup). Criteria for language delay: a. discepancy 1 SD between non-verbal IQ and language performance on the Test of Early Language Delay b. normal cognitive levels and hearing c. mean length of utterance (MLU) at least 2.0 d. no neurological/orofacial abnormality or emotional/autistic problems.	Follow-up at age 7 (after 17–47 months). Measures used at retest: a. PPVT-R b. Token Test c. Photo Articulation Test d. language sample, giving MLU and fluency measures e. Stanford-Binet IQ sub-tests.	1. For three subjects, continued 'stuttering-type' dysfluency is associated with a discrepancy between adequate lexical skills and reduced expressive morpho-syntactic skills. 2. For two subjects, 'normal-type' dysfluency remains and is associated with adequate expressive skills but impaired receptive skills.	4	9	13	Data here selected from nine subjects; one excluded for cognitive delay and three because they received treatment. Focus of study is on language as related to fluency changes over time. If a standard score of 85 on any of the retest measures is taken as a language delay, then all five subjects continue to be language delayed (three showing morpho-syntactic deficits and two showing receptive deficits). However, author not explicit on this point. Measures at follow-up: norm-referenced and criterion-referenced.
Rescorla and Schwartz, 1990 USA	n = 25 all boys Initial age 24–31 months, mean 26.3 months. Recuited via notices in paediatrician offices and local newspaper adverts. Criteria for Specific Expressive Language Delay: a. IQ at least 85 (Bayley Mental Developmental Scale) b. RDLS receptive score in normal range – not more than 4 months' lag to CA c. RDLS expressive score more than 6 months' delay relative to CA. Naturalistic observation and parent report also confirmed expressive delay.	Follow-up after 8–20 months; 16 at age 3 years, 7 at 42/43 months and 2 at 4 years. Measures: 30-min language sample from free play setting; MLU and Index of Productive Syntax (IPSyn) scores reflect length of utterance and syntactic complexity. Criteria for persisting delay: MLU z score ≤ –2 **or** IPSyn z score ≤ –2.	1. 48% still showed severe expressive delay (by MLU z score). 2. 60% still showed severe expressive delay (by IPSyn z score). 3. Higher age at intake (above 26 months) corresponded to worse expressive syntax outcome. 4. Size of expressive lag at age 2 years relative to CA significantly correlated to expressive outcome at 3 years. 5. No correlation between receptive skills at intake and expressive language outcome.	9	11	20	RDLS expressive scale also administered at 3 years of age, but not fully reported. Naturalistic language measures chosen for the follow-up outcomes. 14 of 25 subjects were from families with history of language problems. Role of other factors (such as parental conversational style, motivation to communicate, otitis media history, speech therapy history, articulation deficits) mentioned but not explored. Even the children with better MLU scores (longer utterances) still had poor syntactic/morphological skills. Norm-referenced measures (using Scarborough's norms).

continued

TABLE 29 contd *Natural history studies*

Study	Population sampled	Follow-up duration and measures	Study conclusions	Ranking			Comments
	Subjects; recruitment; criteria for language delay	**Duration; measures; age at follow-up**	**Numbers of children remaining delayed; other observations**	**Replic-ability (of 12)**	**Validity (of 19)**	**Total (of 31)**	
Richman *et al*, 1982 with Stevenson and Richman, 1976 UK	n = 22. Control group of 22 subjects matched for gender, behaviour rating, social class, maternal mental state (but not for IQ). Initial age 3 years. Identified within a population originally sampled at random. Criteria for expressive language delay (mild): RDLS expressive scale score more than 6 months below CA. 11 subjects (50%) had mental age less than 2 years, i.e. non-verbal delay also.	1 year follow-up of language, 5 year follow-up of behaviour, educational status and cognition. Measures at age 4 years: a. language: RDLS expressive scale and English Picture Vocabulary Test (EPVT) b. IQ: Griffiths Scale sub-tests and WPPSI c. behaviour: rating scale. Measures at age 8 years: a. IQ: Weschler Intelligence Scale for Children (WISC) b. behaviour: rating c. education: Neale analysis of reading ability and Schonell spelling tasks.	1. 65% of those language delayed at 3 years still showed persisting delay at 4 years (criterion: RDLS ≤ 40 months at CA 48 months). 2. At age 8 years, 36% sample showed cognitive deficits (criterion: WISC ≤ 85). 3. At age 8 years, 41–45% achieving below 6.5-year level on reading accuracy and comprehension; 50% below 6.5-year level on spelling tasks.	8	15	23	Testers blind to earlier status of children. Language-delayed group had higher than sample representation of behaviour problems, and social/family disadvantages. Norm-referenced measures.
Scarborough and Dobrich, 1990 USA	n = 4 late talkers (LTs); control group of n = 12. LTs: 3 boys, 1 girl. Initial age 30 months. Recruitment – not stated. Criteria for speech/language delay: a. no regular word combinations at 30 months b. normal IQ (General Cognitive Index from McCarthy Scales) c. no gross neurological, hearing or vision impairment. Initial deficits described as 'broad and severe.'	Follow-up duration 5.5 years; language measures up to age 5 and reading, IQ at age 8 years. Measures: a. expressive language (to age 5): syntactic complexity, lexical diversity, pronounciation accuracy, MLU b. receptive language: PPVT (to 8 years) and NSST (to 5 years). c. speech skills: phoneme discrimination/segmentation tasks (to 5 years) and repetition task (8 years) d. IQ: General Cognitive Index (to 5 years) and WISC-R (to 8 years) e. reading: cluster scores on Woodcock Johnson Psycho-educational battery (8 years).	1. Expressive language resolved for all cases by age 5 years, while receptive language did not normalise: a. expressive language composite mean z scores: –3.3 (30 months) –0.2 (5 years) b. receptive language composite mean z scores: –0.7 (30/42 months) –1.3 (4/5 years) 2. Over the period 30–60 months, language deficits appeared to become more selective. 3. Severe reading disability in 3 of 4 LT subjects noted age 8 years.	7	13	20	Explicit statement that subjects were untreated. Author notes mixed results: apparent recovery of language skills in presence of persisting deficits affecting reading development. Proposes model of normal language development in which plateaux of rate of change occur, allowing language-delayed children to catch up for a season. If different language skills hit these plateaux at different ages, then a selectivity of deficits in LTs would appear. Mixed norm-referenced and criterion-referenced measures.

continued

TABLE 29 contd *Natural history studies*

Study	Population sampled	Follow-up duration and measures	Study conclusions	Ranking			Comments
	Subjects; recruitment; criteria for language delay	**Duration; measures; age at follow-up**	**Numbers of children remaining delayed; other observations**	**Replic-ability (of 12)**	**Validity (of 19)**	**Total (of 31)**	
Silva *et al,* 1983 [Dunedin Multi-disciplinary Child Development Study] with Silva, 1980, Silva *et al,* 1987, Silva *et al,* 1982, McGee and Silva, 1982 New Zealand	n = 1027; final sample size 891. Initial age 3 years. Recruitment: population sample – all traceable children born within a 12-month period at a New Zealand hospital. Criteria for language delay: a. general language delay: at or below 5th centile on **both** expressive and receptive task b. specific language delay: at or below 5th centile on one of an expressive **or** receptive task.	Cohort followed 8–11 years. Language measures at ages 3, 5 and 7 years. Additional measures at 7, 9 11 years. Measures: a. language: RDLS expressive and receptive language (to 5 years); ITPA sub-tests (at 7 years) b. IQ: WISC-R c. reading: Burt Word Reading Test d. behaviour: Rutter Parent and Teacher Scales.	1. Prevalence of any type of language delay: age 3: 7.6% age 5: 10.4% age 7: 8.4%. These cases not same children but a fluctuating group. 2. 124 of 168 children (74%) evidenced a type of language delay at one assessment only. 3. General language delay more stable than a specific language delay (78.2%). 4. General language delay more often associated with a reduced IQ measure. 5. Specific expressive delay age 3 associated with a language delay at ages 5 and 7. 6. Specific receptive delay age 3 not associated with later language delay, but 45.8% went on to have reduced IQ/reading measures at age 7.	6	17	23	Sample slightly disadvantaged in terms of SES relative to New Zealand figures. Sample had under-representation of Maori/Polynesian groups relative to New Zealand as a whole. 'Normalisation' of delayed cases not explicit. Norm-referenced measures.
Thal and Tobias, 1992, with Thal and Bates, 1988, Thal *et al,* 1991 USA	n = 10 LTs, also controls; 10 language matched and 10 age matched. Initial age (LT) 18–29 months, mean 22 months. Recruitment: families self-referred after local newspaper adverts. Criteria for LT: a. 0–64 single words expressively; no word combinations b. at or below 10th centile on Language and Gesture Inventory c. no history hearing loss d. no repeated ear infections e. no mental retardation f. no behavioural/neurological impairment.	1 year follow-up. Measures: a. from a language sample: word count; MLU; Brown's Stage Level (of syntactic development). Additional baseline measures: a. receptive skills (experimental forced choice task) b. communicative gestures c. Language and Gesture Inventory.	1. Identified 40% as truly delayed and 60% as LTs. 2. LTs showed mean productive vocabulary increase of 525 words, compared with 300–350-word increase expected in normal-language development.	6	9	15	State that no referral of LT group made for speech/ language assessment. Larger sample size replication underway. Norms for development of communicative gestures not yet established. Risk factors for true language delay: delayed receptive language and delayed gesture production. Positive signs (with LTs): age appropriate comprehension at outset, and compensatory use of gesture. Criterion-referenced and norm-referenced measures.

TABLE 30 *Natural history evidence from predictive validity studies*

Study	Population sampled	Follow-up duration and measures	Study conclusions	Ranking			Comments
	Subjects; recruitment; criteria for language delay	Duration; measures; age at follow-up	Numbers of children remaining delayed; other observations	Replicability (of 12)	Validity (of 19)	Total (of 31)	
Klee *et al*, 1997 USA	Subsample of n = 6 untreated cases. Initial age 2 years. Original sample recruited via birth announcements, notices in physicians' offices. Criteria for delay: clinical concern plus more than 1 SD below mean on receptive/expressive language measures.	1 year follow-up to age 3. Measures: a. audiological screen b. Infant Mullen Scales of Early Learning (MSEL) (age 2) or pre-school (age 3) c. measures of MLU and word count from a language sample.	1. Four of six cases (67%) persisted from age 2 to age 3 years. 2. Two further cases considered within normal limits age 2 were of clinical concern age 3 years.	2	13	15	Sample data from larger study (n = 36) of predictive validity of a screen at age 2 years. Individual subject data, and treatment status specified. Norm-referenced measures.
Renfrew and Geary, 1973 UK	n = 150. Initial age about 5 years (school entry). Recruitment: children in schools in Oxford area. Criteria for speech delay: failing an articulation test.	6 month follow-up. Measures: a. non-standardised speech (articulation) tasks b. RAPT of expressive language.	1. 54% of the 150 children still had incorrect articulation. 2. Of a subgroup of 76 children with 'serious' speech problems, 51% still had incorrect articulation. 3. Within the speech delay group, 46 originally evidenced language delay. 74% had resolved this language delay. 4. The fewer the original number of errors, the better the spontaneous recovery.	4	15	19	Children who had speech therapy omitted from these figures by the author. Predictors of spontaneously resolving speech delay: a. ability to imitate target speech sounds b. ability to discriminate like-sounding words c. ability to say tongue twisters d. (to lesser degree) auditory memory skills. Criterion-referenced measure.
Ward, 1992	Groups 1, 2, 3: Group 1; n = 119 mean age 12.2 months Group 2; n = 61 mean age 10.5 months Group 3; n = 23 mean age 9.7 months. Recruitment: all infants brought during 1 year for a routine hearing test. Criteria for delay: Failing expressive and/or receptive sections of Ward Screening test; plus expressive and/or receptive delay on Receptive Expressive Emergent Language (REEL) Scale.	1 year follow-up Of remaining 199 subjects, 18 excluded here as being secondary language delays (associated with developmental delay, hearing loss or emotional disorder). Measures at follow-up: REEL.	1. Group 1: Expressive/ receptive delay and listening difficulties: 81 of 99 cases (82%) persist; six cases now specific expressive delay. 2. Group 2: Expressive/ receptive delay, no listening difficulties: 44 of 60 cases (73%) persist; seven cases now specific expressive delay. 3. Group 3: Expressive delay: 11 of 22 cases (50%) persist; six cases now expressive/ receptive delay.	6	17	23	Type of language delay changes over time. A greater percentage of Groups 1 and 2 continued to show language delay than Group 3.

TABLE 31 *Intervention studies – RCT designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Almost and Rosenbaum, 1998	n = 26 (13E, mean age 42 months, SD 7.9; 13C, mean age 41 months, SD 7.3, with no age range provided). A further two subjects from each group were lost due to refusal to comply with group allocation or failure to return to the clinic. Participants were selected from those referred to a speech and language clinic for assessment and intervention. The sample was 81% male, with severe, specific phonological disorders on the Assessment of Phonological Processes – Revised (APP-R) (Hodson, 1986). Hearing, receptive language on the RDLS (Reynell, 1977), and oral structures and function were all within normal limits.	RCT cross-over design. Individual subjects were randomly allocated to either a treatment group (receiving speech therapy) or a delayed-treatment control group. All assessments were 'blinded'.	Articulation/ phonology	The study was designed to determine the effectiveness of speech therapy for children with phonological disorders. Direct treatment was provided by the first author in a hospital speech and language department. Treated subjects were seen twice-weekly for 30 min over a 4-month period (i.e. 22 sessions on average). Four to six targeted phonological processes were treated initially by means of minimal pairs contrasts and later by means of correct production in conversation within a modified cycles approach (Monahan, 1984; Hodson and Paden, 1983). Following completion of treatment for the initial group the delayed therapy group were provided with treatment while the first group did not receive any intervention.	After the first 4 months of intervention treated children made significantly greater gains than controls ($p < 0.01$) on the standardised GFTA (Goldman and Fristoe, 1969) and the following criterion-referenced measures: the APP-R (Hodson, 1986) and the Percentage Consonants Correct (PCC) (Shriberg and Kwiatkowski, 1982), revealing progress in both single word and conversation contexts. The groups did not differ pre- or post-intervention on MLU (Brown, 1973), a general measure of expressive language. The delayed treatment group's response to the second period of intervention was similar to that of the initial treatment group with one exception: the initial treatment group made more progress on the PCC (a measure of precision in articulation in conversation contexts) ($p < 0.05$). **Comments** The results confirm the effectiveness of standard treatment approaches in a community setting. However, the phonological treatment programmes did not generalise to expressive language, though not all of the children had marked problems in expressive language: many had MLUs within the normal range (see Miller, 1981) prior to intervention. Results of an analysis including the children lost to follow-up were also reported. Statistical power analysis was used to determine sample size (see Cohen, 1992). However, analysis was carried out on raw scores of GFTA, not on standardised scores, thus losing control over the effects of maturation and losing information about the normalisation of scores in the experimental and control groups.
Fey, Cleave, Long and Hughes, 1993	n = 30 (11E1, age range 47–67 months, mean 55 months, SD 6.05; 10E2, age range 44–70 months, mean 56 months, SD 7.21; 9C, age range 45–67 months, mean 56 months, SD 6.31). Most of the children were referred by their parents in response to media coverage and the rest were referred by speech-language pathologists. The children's scores on the Developmental Sentence Score (DSS) (Lee, 1974) were < 10th percentile (referenced to the lower of CA or mental age), indicating delay in expressive grammar, and all had non-verbal IQs of > 70 on the Leiter International Performance Scale (LIPS) (Leiter, 1979). However, there were marked differences within the groups in IQ, phonology and language comprehension. 70% of the overall sample was male. The authors note that the sample should not be considered as having specific language impairment.	RCT design. Individual subjects were randomly allocated to either a treatment group (clinician-administered or parent-administered) or to a control group. All pre-testing was 'blinded'.	Expressive language (syntax)	The study was designed to compare the effectiveness of clinician-administered treatment of expressive syntax (grammar) delay with that administered by parents. Children in the clinician group were seen in a clinic setting by a speech-language pathologist for three 60-min sessions per week for a period of 20 weeks. The children were seen individually one per week and twice per week in a group of 4–6 children. Four goals were identified for each child. Treatment consisted of imitation of the target and of a contrast form and focused stimulation procedures (e.g. modelling and recasting of sentences under naturalistic conditions). Cyclical goal-attack strategies were used, with one goal targeted each week. Targets were either dropped or combined when they were used productively, thus making space for the introduction of the next goal. Treatment in the parent group consisted of: (a) for the first 12 weeks, a weekly 2-hour meeting between the therapist and 4–6 parents in the clinic (with no children present) and three 60-min home visits by the therapist; (b) a monthly 60-min meeting in the clinic for the remaining 8 weeks to informally check on treatment fidelity.	There was no difference ($p = 0.47$) in outcomes between the clinician-treated and the parent-adminstered groups. Children in both treatment groups made significantly greater gains than the controls on overall DSS scores (developmental measures of expressive syntax which can be interpreted in normative terms) obtained from analyses of pre- and post-intervention language samples of parent–child interaction ($p = 0.0002$), and also on the mean verb score and for the percentage of sentences awarded a sentence point (Lee, 1974). There was no difference between treated and untreated children for the personal pronoun measures ($p = 0.64$) as most of the children used pronouns effectively prior to intervention. **Comments** The results indicate the effectiveness of focused stimulation procedures for treating expressive syntax problems, with parent-administered treatment resulting in gains as large as those from direct, intensive treatment by a clinician. The implications for cost-effectiveness of treatment (126 vs. 240 hours of clinician's time, in favour of the parent treatment) are noted. However, the control group failed to make any gains due to maturation over the 4.5-month interval, which increased the treatment effects. Also, children in the clinician group made more consistent gains due to greater variability in the gains in the parent group. Lastly, the sub-group which made the greatest gains included parents who tended to produce the fewest recasts and who changed the least following treatment, which raises the possibility that the observed gains resulted from some factor other than the focused stimulation procedures. The authors' checks on reliability of coding yielded agreement in the range 80.6–100%.

* *E = experimental; C = control*

continued

TABLE 31 contd *Intervention studies – RCT designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Gibbard, 1994 (Study 1)	n = 36 (18E, age range 29–39 months, mean 35 months, and 18C, range 27–39 months, mean 32 months). Children were selected from a normal clinical population at a health centre following referral by a health visitor. The mothers of two further children invited to participate declined. Children had a vocabulary of fewer than 30 words; passed all sub-sections on the Denver Developmental Screening Test (DDST) (Frankenburg, Dodds and Fandal, 1973) apart from language development; had no medical condition suggestive of a language delay including history of otitis media; and had not previously received any speech and language therapy. Sixty-nine per cent of the sample were boys, and 66% had non-manual SES.	RCT design. Pairs of individual subjects were matched for gender, birth order, SES and age and one from each pair was randomly allocated either to the treatment group (parent-administered intervention) or to the control group. The study was carried out over two phases.	Expressive language (and receptive language)	The study was designed to evaluate the effectiveness of parent-based intervention with regard to increasing the expressive language skills of children with language difficulties from a single-word level to a three- to four-word level. The mothers attended fortnightly training sessions of 60–75-min in length run by a therapist in a clinic setting for 6 months (11 sessions). At each meeting, parents were given objectives for work with their child, together with suggestions for activities and methods for meeting the objectives (many based upon the Derbyshire Language Scheme (DLS) (Knowles and Masidlover, 1979). Emphasis was placed upon transfer of language skills to everyday situations. Mothers were given the opportunity during the sessions to work in small groups and to identify suitable activities. Structured teaching approaches were used by the therapist to clarify each language objective for the mothers.	The outcome measures were pre- and post-intervention scores on: (a) two standardised tests: the RDLS (Reynell, 1985) and the RAPT (Renfrew, 1986); and (b) criterion-referenced measures obtained from a mother–child language sample (e.g. MLU, a score for all one-word utterances, and a total score based on all utterances); from scores on the DLS Picture Test and from parental report regarding vocabulary used and the structure of the child's utterances. The results revealed that children in the experimental group whose mothers were given the training made significantly greater gains on all measures than the children in the control group ($p < 0.01$). **Comments** While the children in the control group made progress over the 6-month period, the children of the parents who received training made significantly greater progress, thus demonstrating the value of group-based parental language intervention, which is also more cost-effective than direct treatment by a clinician. The effects of the intervention also generalised to the children's language comprehension scores on the RDLS. An analysis of covariance (ANCOVA) was used, with pre-test scores as covariate, despite the fact that there were no significant differences between pre-test scores. In addition, analyses appear to have been carried out on the raw scores of the RDLS and RAPT, thus losing control over the effects of maturation that standardised scores afford, and losing information about the normalisation of scores in the experimental and control groups. The authors' checks on reliability of coding yielded agreement in the range 93–96%.
Girolametto, Pearce and Weitzman, 1995	n = 16 (8E, age range 22–38 months, mean 30 months, SD 4.9; 8C, age range 23–34 months, mean 28 months, SD 4.7). Participants were recruited from waiting lists for parent-focused intervention. The children were all aged between 23–33 months, were at the single-word stage of language acquisition and had delayed vocabulary acquisition on the McArthur Communication Development Inventories (CDI) (Fenson *et al,* 1993), scored within +2 SDs of the mean of the Vineland Adaptive Behaviour Scales (Sparrow, Balla and Cicchetti, 1984), had normal hearing, no oral-motor problems, no neurological problems and did not have autism or other persistent developmental disorder. English was the sole language of the home. The children all had delayed expressive language skills (within the range 6–24 months) and poor vocabularies (< 5th percentile on the CDI). Five children in each group also had a receptive language delay of more than 6 months. All of the mothers had completed high school. 69% of the overall sample were boys.	RCT design. Individual subjects were randomly allocated to either a treatment group (parent-administered intervention) or to a control group.	Expressive language (vocabulary)	The study was designed to examine the effectiveness of training parents to use focused stimulation techniques to help the development of their children's vocabulary. Two speech/language pathologists and a trained parent associate delivered the Hanen Program for Parents consisting of seven evening sessions of unreported length which used lectures, role play, focused discussions and videotapes to teach parents strategies consistent with a focused stimulation approach and three individual consultations over a 10-week period. Parent–child interactions were video-taped during the consultations and were used to provide parents with individual feedback. Target and control vocabulary items that were understood but not produced expressively by the child were identified for each individual child. Objects and photographs representative of the words served as two semi-structured probes to assess the child's production of the items. Overall attendance was judged to be satisfactory and all mothers reported use of specific programme techniques. The control group had a significantly higher mean pre-test score on the Child Behaviour Checklist (CBCL) (Achenbach, 1986) Externalising Score ($p < 0.04$) but there were no other pre-test differences. All assessments were 'blind'.	Criterion-referenced measures of expressive vocabulary derived from the use of the two semi-structured probes (e.g. the number of different target words and different control words used across the probes) were used, together with parent reports on expressive vocabulary and use of symbolic play gestures from the CDI and scores on the Internalising and Externalising Scales of the CBCL. The results revealed that the experimental group children produced significantly more target words on average during the semi-structured probes than the controls ($p < 0.02$). However, there was no difference between the two groups in production of control words or reported vocabulary size from the CDI. The experimental group also increased their use of symbolic gestures compared with the controls ($p < 0.03$). Scores on the CBCL revealed a significant reduction ($p < 0.02$) in the experimental children's externalising scores relating to aggressive and destructive behaviour following the parent training but no change in the control group's score and no differences between the two groups in internalising or total scores for the CBCL. **Comments** The results show that parents can be trained to use focused stimulation to teach their children vocabulary. The children were also able to generalise their production of vocabulary to the probe tasks which involved unfamiliar adults, settings and objects. The perceived improvement in the ratings of the experimental group's behaviour following parent training is also interesting, though it is unclear in the absence of direct observation whether the change was due to an actual improvement in behaviour or to a change in parental expectations. Analyses of CDI data were carried out on raw scores rather than standardised percentile scores which would have provided information about the degree of normalisation in the treatment and control groups over the 10-week pre-test–post-test period. The authors' checks on reliability of coding of semi-structured probe data yielded average agreement of 92%.

* *E = experimental; C = control*

continued

TABLE 31 contd *Intervention studies – RCT designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Girolametto, Pearce and Weitzman, 1996	n = 25 (12E, age range 25–35 months, mean 29 months, SD 3, and 13C, age range 23–34 months, mean 29 months, SD 3), overall age range 23–35 months). Participants were all on waiting lists for parent-focused language intervention and were referred by their parents in response to notices in the press and in health and educational settings. The children were all aged between 23–33 months, had normal hearing, no oral-motor problems, no neurological problems and did not have autism or other persistent developmental disorder. English was the sole language of the home in all cases. There was marked variability in the children's receptive language skills, phonological skills and cognitive abilities. However, only four children in the experimental group and one in the control group had a receptive language delay of more than 6 months on the Sequenced Inventory of Communication Development (SICD) (Hendrick *et al,* 1984). Similarly, all but one child in the experimental group and two in the control group had IQs < 85 on the Stanford Binet (Thorndike *et al,* 1986) or Developmental Profile II (Alpern *et al,* 1984) indicating cognitive abilities within the normal range for most subjects. All of the families were middle-class (0% non-manual) but no information about the gender balance in the sample was provided. None of the control group received intervention prior to the start of the study but around four received community-based treatment as the study progressed.	RCT design. Individual subjects were randomly allocated to either a treatment group (parent-administered intervention) or to a control group.	Expressive language (vocabulary and syntax)	The study was designed to examine the effectiveness of training parents in focused stimulation techniques to help the development of their children's expressive language skills. Two experienced speech/language pathologists and a trained parent associate delivered the Hanen Program for Parents. The 11-week programme consisted of 12 training sessions (no details provided), eight evening sessions (each lasting 150 min and designed to teach parents strategies using lectures, role play, focused discussions and videotapes) and three home visits. During the home visits parent–child interactions in free play were videotaped and viewed to provide parents with individual feedback. Twenty vocabulary items that were understood but not produced expressively by the child were identified for each child by means of assessment and parent report. Ten served as target words and the rest as controls. Treatment fidelity was determined by attendance at the training sessions, group meetings and home visits together with evidence from parent–child interactions that target words were being used by mothers. Overall attendance was judged to be satisfactory and all mothers were observed using targets. Pre-tests revealed no significant differences between the two groups of children in terms of age, IQ, mother's age and years of schooling or on any of the dependent variables used in the study. All pre- and post-intervention assessments were 'blinded'.	The outcome measures were: criterion-referenced measures of maternal interaction and of the children's performance in the free play sessions at home; and the number of target and control words produced by the children during the probes. The results revealed that the trained mothers increased their use of language modelling techniques and as a result reduced their MLU and their rate of talking compared with control mothers ($p < 0.01$). The experimental mothers also used a greater number of target words with their children ($p < 0.01$). The treated children increased the size of their vocabulary on the CDI more than the children in the control group ($p < 0.01$) and also used a greater number of words in the free-play settings with their mothers ($p < 0.01$). There were no differences between the two groups in terms of number of utterances or rate of words per min, but the parents of the treated children reported that the structure of the children's utterances was more mature ($p < 0.01$) and the treated children used more multiword combinations at post-test ($p < 0.05$). The treated group also produced significantly more of the target words in response to post-test probes and used a greater diversity of targets ($p < 0.02$). They also used more of the control words than the untreated group ($p < 0.05$). **Comments** The results reveal that the trained mothers used the focused stimulation approaches and that their children's expressive language skills improved more than the effects of maturation as a result. However, as the authors note, the mothers here were well-educated and highly motivated and may not be representative. Multivariate ANOVA was used to analyse the data on account of the multiple measures used. The number of vocabulary words on the CDI was reported but the authors did not use the percentile scores from the CDI norms which would have provided information about the degree of normalisation in the treatment and control groups over the 4-month pre-test–post-test period. The authors' checks on reliability of coding yielded agreement in the range 72–100%.

* *E = experimental; C = control*

continued

TABLE 31 contd *Intervention studies – RCT designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Lancaster, 1991	n = 15 (5E1; 5E2; 5C; overall age range 40–53 months, mean 43 months, SD 3.2) Participants were selected from those referred to a speech and language clinic for assessment and intervention. The sample was 80% male, with specific phonological disorders, i.e. standard scores on the EAT (Anthony, Bogle, Ingram and McIssac, 1971) < 80, and hearing, receptive and expressive language all within normal limits on the RDLS (Reynell, 1977). The children had no physical handicaps and English was the main language of the home.	RCT design. Individual subjects were randomly allocated to either a treatment group (clinician-administered or parent-administered) or to a control group.	Phonology	Comparison between a control group and two experimental groups: (a) a group receiving direct treatment from a speech and language therapist in a clinic setting based upon 'eclectic principles' centred around minimal contrast therapy, help with production of individual skills, work on metalinguistic skills and practice in producing target sounds. Treated subjects were seen weekly or fortnightly for 30–40 min per session over a 6-month period, with an average of 7.6 contact hours; (b) a group whose parents received training in delivering a programme based on an 'input' approach involving target words and sounds in 6-week cycles (Hodson and Paden, 1983) with no direct therapy. Length and frequency of parents' sessions with the children is not given but the therapists' involvement in training and monitoring the parents was no greater than 4 hours per child over the 6-month intervention period (i.e. around 50% of time for direct therapy).	Treated children made significantly greater gains ($p < 0.05$) than the children in the control group on the CDS (Hodson and Paden, 1983), a criterion-referenced measure of severity of phonological disorder, with an age-adjustment for older children with CA > 48 months. There was no significant difference between the progress made by the direct treatment and parent treatment groups ($p > 0.1$). **Comments** The results reveal the effectiveness of parent treatment but the small number of children in each group results in a lack of sensitivity in comparisons, particularly between the two treatment groups (Cohen, 1992), where the trend was for the direct treatment group to make greater gains. However, three to five children in both groups showed some degree of normalisation of scores compared with none in the control group.
Methany and Panagos, 1978	n = 24 (8E1; 8E2; 8C; overall age range 65–82 months, no mean provided). The subjects presented with articulatory and syntactic problems and all made at least seven consonant errors on the Photo Articulation Test (Pendergast, Dickey, Selmar and Soder, 1984) and were able to achieve no higher than Programme Number 6 on the Programmed Conditioning Language Test (Gray and Ryan, 1973). No information was provided about gender balance but none of the children came from ethnic backgrounds.	RCT design. Individual subjects were randomly allocated to either a treatment group (articulatory training or syntax training) or to a control group.	Articulation and expressive language (syntax)	Comparison between the effects of articulatory training and syntax training upon syntax and articulation skills over a 5-month intervention period. Training in both intervention groups was based upon Monterey Language Programs (Gray and Ryan, 1973; Ryan and Baker, 1971) which are highly-structured and emphasise imitation and reinforcement for correct production of language forms. One intervention group was trained on a syntax programme while the other trained on an articulation programme. No details of frequency or duration of treatment sessions is provided. The control group received no treatment.	Results reveal that the two intervention groups made significantly greater progress on both the Photo Articulation Test and the Programmed Conditioning Language Test norm-referenced tests than the control group ($p < 0.05$). **Comments** Articulation and syntax intervention appeared to be equally effective. Thus syntax intervention indirectly improves articulation and articulatory intervention appears to indirectly improve syntax. Note that the small numbers in each group reduces the sensitivity of comparisons, particularly between the two treatment groups (Cohen, 1992). In addition, statistical analysis was carried out on raw scores rather than normative scores.

* *E = experimental; C = control*

continued

TABLE 31 contd *Intervention studies – RCT designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Reid, Donaldson, Howell, Dean and Grieve, 1996	n = 30 (8E1; 7E2; 8C1; 7C2; the overall age range was 42–66 months but no information was provided concerning either SD, mean age or gender balance). Participants were selected from those with specific phonological delay and had normal hearing, scores within normal limits on the RDLS for Verbal Comprehension (Reynell, 1985) and standard scores of < 85 on the EAT (Anthony *et al,* 1971).	RCT design. Individual subjects were randomly allocated to either a treatment group (Phase 1 of the Metaphon programme [Dean *et al,* 1990] only, or Phase 1 plus Phase 2) or to an associated control group whose progress was monitored over either a 6-week period or a 10-week period.	Phonology, phoneme awareness and receptive vocabulary	A preliminary study comparing the effectiveness of Phase 1 of the Metaphon (one weekly 30-min direct treatment session for 6 weeks focusing upon games and activities to help children learn how to detect and classify sounds in words) with Phase 1 plus Phase 2 (which involves a further 4 weekly sessions using minimal pairs to develop communicative awareness and effectiveness) in a clinic setting.	Scores on the EAT showed a significant increase ($p < 0.02$) only for the Phase1 plus Phase 2 group. The 10-week intervention group also made significant gains on a criterion-referenced test of phoneme awareness ($p < 0.03$) though not on tests of rhyme awareness, word order awareness nor on the standardised BPVS (Dunn *et al,* 1982). The Phase 1 and control groups did not make significant pre-test–post-test gains. **Comments** Results for each of the four groups are reported only as gains within each group: i.e. there is no direct comparison between gains for Phase 1 and those for Phase 1 plus Phase 2 nor for the two treatment groups and their associated control group. Results from the control groups reveal significant spontaneous improvements in phonological processes in the absence of treatment and highlight variability in performance and the effects of maturation over even a 6-week test–re-test interval. However, no SDs are reported and hence no effect sizes can be calculated.
Schwartz, Chapman, Terrell, Prelock and Rowan, 1985	n = 10 (8E, age range 32–40 months, mean 36 months, SD 2.9; 2C, age range 34–35 months, mean 34 months, SD 0.71). Participants were ten children, with normal hearing and no physical, sensory or motor deficits or history of emotional disturbance. They were of average cognitive ability (non-verbal IQ of > 85 on the LIPS) (Arthur, 1952) but had delays in language comprehension of 6–9 months below mental age on the TACL (Carrow, 1973) or the Auditory Comprehension sub-test of the Preschool Language Scale (PLS) (Zimmerman *et al,* 1969) and delays of 12–16 months below mental age on expressive language on the DSS (Lee, 1974) and the Verbal Ability sub-test of the PLS (Zimmerman *et al,* 1969). All of the children were boys (gender balance 100%). Seven of the children were receiving regular speech and language therapy two to three times per week during the course of the study.	RCT design. Two subjects were randomly allocated to a control group and the rest took part in the clinician-administered treatment group.	Expressive language (syntax)	The aim of the study was to examine the appropriateness of an intervention programme using vertical discourse structures observed in normal development of two-word utterances (Scollon, 1976) for increasing the use of multiword utterances by children with specific language impairments. The programme used picture and object stimuli which demanded two-word utterances, and adult prompts to request clarification and encourage the child to produce the first term, and then the second, i.e. successive single-word utterances. These were then 'recast' by the adult to facilitate the emergence of two-word and multiword utterances. Prior to the experimental and control sessions the children were given a pre-test and scores obtained using 24 individualised stimuli. Following this, two children were randomly assigned to a control group which entailed participation in another unrelated study. Children in the experimental group were seen three times a week for ten sessions over a 3–4-week period (the length of the sessions was not given) and were presented with 16 experimental stimuli (eight picture and eight enacted by the experimenter using objects and dolls) and 16 unrelated non-experimental stimuli. Experimental stimuli were selected to represent noun–noun constructions which encoded four given semantic relations and were presented in random order.	The outcome measures were pre-test and post-test scores for combined successive single-word utterances and multiword combinations collapsed across sessions. The results revealed that the experimental group significantly increased their scores ($p < 0.05$) while the scores of one of the children in the control group fell. Six of the experimental group recorded their largest number of multiword utterances during the last three sessions. The authors conclude that the vertical structure procedure employed increased the number of word combinations produced by the children in the experimental group. **Comments** The small number of children involved in this study, particularly in the control group, poses problems for generalisation of the findings. Such small numbers do not provide adequate control for the effects of maturation given the observed variability in scores and the fact that multiword utterances were beginning to be produced in the pre-test. In addition, the rationale for combining the successive single-word utterances and multiword utterances to form a single composite score was unclear, particularly as the production of multiword utterances seemed markedly greater following intervention than single-word utterances and given the prediction of an increase in multiword combinations. The authors' checks on reliability of coding yielded agreement in the range 80–99%.

* *E = experimental; C = control*

continued

TABLE 31 contd *Intervention studies – RCT designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Shelton, Johnson, Ruscello and Arndt, 1978 (Study 1)	n = 45 (15E1; mean age 47 months, SD 6; 15E2, mean age 49 months, SD 6; 15C mean age 39 months, SD 12). No age ranges were provided. Participants were children with normal hearing and were matched across the three groups in terms of receptive vocabulary scores and bilingual backgrounds (two children in each group), with below-average scores for their age on the Templin-Darley Articulation Test (Templin and Darley, 1969). Sixty children were originally included but five dropped out leading to a further loss of ten matched children. No information about gender balance was provided.	RCT design. Sets of three children were matched on receptive vocabulary scores and bilingual background and randomly allocated to either a treatment group (parent-administered listening training or parent-administered training based upon reading and talking) or to the non-intervention control group.	Auditory listening skills; articulation and receptive language (syntax)	Comparison between the effects of two different approaches to home-based parent-administered listening training for children with articulatory difficulties to improve correct target sound identification. Training was carried out at home by parents in both listening (5 min per day, 5 days per week for 57 sessions) and reading/ talking (15 min per session, 5 days per week for 57 sessions) treatment groups was based upon four target sounds with 14 lessons per sound. Fortnightly telephone calls and end of intervention interviews were used to check on treatment fidelity.	Outcomes for six norm-referenced tests (the Goldman-Fristoe-Woodcock Test of Auditory Discrimination (Goldman, Fristoe and Woodcock, 1970); the NSST (Lee, 1971); the Auditory Association Test from the ITPA (Kirk and Kirk, 1968), Templin-Darley Articulation Screening Test (Templin and Darley, 1969); and the Deep Screening Test of Articulation (McDonald, 1964)) and two criterion-referenced measures of auditory discrimination reveal only one difference ($p < 0.05$) between the three groups: the control group made greater gains on the noise condition of the Goldman-Fristoe-Woodcock Test. **Comments** The results reveal no significant treatment effects ($p > 0.1$) for either of the two listening training approaches. Parent-administered listening training did not improve either listening skills or articulation. There were no direct observational checks on treatment fidelity. Analyses were carried out on post-test–pre-test gain scores and on raw scores for standardised tests.

* *E = experimental; C = control*

TABLE 32 *Intervention studies – quasi-experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Conant, Budoff, Hecht and Morse, 1984	n = 48 (26E, age range 36–69 months, mean 53 months SD 8.7; 22C, age range 32–72 months, mean 51 months, SD 9.6). Children participating were aged between 3–8 years, had normal hearing and vision but had a MLU between 1.0–2.0 and serious and pervasive language disabilities. Excluded were autistic children, children with individual language remediation programmes and those with sensory impairments. No information about SES or gender of the participants were reported. The sample contained a wide range of cognitive abilities. Twenty-two had secondary language delay (i.e associated with serious delay in cognitive functioning 26 children (9E and 13C) had primary language delay and it is these children who are considered here.	Quasi-experimental design. Children were assigned non-randomly to either the treatment group (training in pragmatic skills) or to the non-intervention contrast group.	Pragmatics and expressive language (syntax)	This study investigates whether it is possible to teach the appropriate and effective pragmatic uses of language in conversational settings while at the same time teaching the comprehension and production of specific syntactic forms. The intervention approach used a series of communication games which provided goals, feedback, a meaningful communicative context, opportunities for practice, incidental learning and adult and peer models of effective language use. The target children played communication games 2–3 times per week for 30–45 min on each occasion for a period of 4 months.	The outcome measures were four composite values (amount of speech, length of units, syntax and speech acts) constructed from a range of criterion-referenced measures. The data were analysed using a multivariate ANCOVA, in view of significant pre-test differences between the treatment and contrast groups. With regard to the children with primary language delay, the results revealed that the treatment group achieved significantly higher scores than the contrast group on all four language measures ($p < 0.035$ overall, and $p < 0.05$ for individual measures). Interestingly, children with secondary delays did not make progress in response to the treatment. **Comments** The results here suggest that the use of the communication games approach improved not only the children's pragmatic use of language but also the structure of their expressive language. However, the intervention was only successful for children with primary language delay. Information about the degree of normalisation which occurred over the intervention period would have been useful. The author's checks on reliability of coding yielded levels of agreement between 83–99%.

* *E = experimental; C = control*

continued

TABLE 32 contd *Intervention studies – quasi-experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Fey, Cleave, Ravida, Long, Dejmal and Easton, 1994	n = 26 (10E1, 8E2 and 8C, a delayed-treatment control group). The mean age of the subjects was 56 months (range 44–70 months, SD 6.3). Participants were selected from a pool of 30 children with primary language delay (non-verbal IQs > 72 and DSS (Lee, 1974) at or below the 10th percentile for the lower of CA or mental age, indicating delays in grammar, and moderate to profound difficulties in phonology). Seventy per cent of the final group of 26 children were boys.	Quasi-experimental design. Subjects were randomly assigned to either a clinician-administered treatment group or to a parent-administered treatment group but were selected non-randomly for immediate treatment or for the delayed treatment control.	Phonology and expressive language (syntax)	This study investigates whether gains in grammar can also lead indirectly to gains in phonological development and also compares the effectiveness of direct clinician treatment with parent treatment. Children assigned to clinician treatment were seen three times per week for 1 hour for a 5-month period. Parents in the parental intervention group received 12 weekly sessions of initial training by a speech and language therapist lasting 2 hours per session. Each child in the parent group also received a monthly individual session with a speech and language therapist. After the initial training, parents continued to receive one monthly individual and one monthly group session designed to provide on-going support. Each child in the study worked through four specific targets in language form in a cyclical format. The primary intervention model used focused stimulation techniques (including modelling of target forms and the use of sentence recasts) in naturalistic tasks. Ten minutes of imitative drill weekly were also incorporated into the clinician programme. Details of how parents were asked to organise their time with the children in the parental treatment condition are not given.	Post-intervention scores on measures of expressive syntax (the DSS) and phonology (the APP-R) (Hodson, 1986) and the PCC (Shriberg and Kwiatkowski, 1982) were analysed. The results revealed highly significant gains in expressive syntax on the DSS ($p < 0.002$). There were no differences in outcome between clinician treatment and parent intervention for expressive syntax, indicating that they were equally effective. However, neither of the treatments had any effect upon PCC post-test scores ($p > 0.1$), though there was considerable individual variation across groups, indicating that the treatment was not effective for phonology. (These findings were replicated by results from the control group after they received treatment.) **Comments** In contrast to the findings of Methany and Panagos (1978) treatment of expressive syntax does not generalise to phonology. However, the children in the present study had more severe phonological disorders than those in the Methany and Panagos sample.

* *E = experimental*

continued

TABLE 32 contd *Intervention studies – quasi-experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Gibbard, 1994 (Study 2)	n = 25 (8E1, age range 27–39 months, mean 32 months; 9E2, age range 29–36 months, mean 32 months; 8C, range 29–35 months, mean 31 months). Children were selected from a normal clinical population at a health centre following referral by a health visitor. The mothers of three further children invited to participate in the study declined. Children had a vocabulary of fewer than 30 words; passed all subsections on the DDST (Frankenburg, Dodds and Fandal, 1973) apart from language development; had no medical condition suggestive of a language delay including history of otitis media; and had not previously received any speech and language therapy. The three groups did not differ in non-verbal ability scores on subtests of the McCarthy Scales of Children's Abilities (McCarthy, 1972). Seventy-six per cent of the sample were male, and 35% had non-manual SES.	Quasi-experimental design. Subjects were assigned to either a treatment group (individual direct treatment group or a parent language programme group) or to a control group. Subjects were matched for gender, age, birth order and SES across the groups. (Mothers were 'invited' to join either a treatment group or the control group, hence allocation to groups appears not to have been random.)	Expressive language (and receptive language)	The study was designed to compare the effectiveness of parent-based intervention for children with expressive language difficulties with direct treatment approaches administered by a clinician. In one treatment group, the children received weekly individual speech and language therapy for 30 min from a clinician over a 6-month period. In the second treatment group, the mothers attended fortnightly training sessions of 60–75 min in length run by a therapist in a clinic setting for 6 months (11 sessions). At each meeting, parents were given objectives for work with their child, together with suggestions for activities and methods for meeting the objectives (many based upon the DLS) (Knowles and Masidlover, 1979). Emphasis was placed upon transfer of language skills to everyday situations. Mothers were given the opportunity during the sessions to work in small groups and to identify suitable activities. Structured teaching approaches were used by the therapist to clarify each language objective for the mothers. The control group met fortnightly over a 6-month period and provided mothers with training in non-specific cognitive activities unrelated to language, thus controlling for possible Hawthorne effects (i.e. that improvements from the parent language group might be the result of non-specific variables unrelated to the language programme).	The outcome measures were pre- and post-intervention scores on: (a) two standardised tests: the RDLS (Reynell, 1985) and the RAPT (Renfrew, 1986); and (b) criterion-referenced measures obtained from a mother–child language sample (e.g. MLU, a score for all one-word utterances, and a total score based on all utterances); from scores on the DLS Picture Test and from parental report regarding vocabulary used and the structure of the child's utterances. The results revealed that children in the parent language group made significantly greater gains on all measures compared with those in the control group ($p < 0.05$), demonstrating the efficacy of the specific language intervention. However, the individual treatment group made significant gains compared with the control group in only the total scores from the DLS and MLU. The only significant difference between the children in the parent language group and those receiving individual direct treatment was in the larger MLUs found in the case of the children in the parent-administered group ($p < 0.01$). **Comments** While the children in the control group made progress over the 6-month period, the children of the parents who received training made significantly greater progress, thus demonstrating the value of group-based parental language intervention. The effectiveness of the parent language group programme appeared to be at least as effective as individual, direct treatment and possibly more consistent in its outcomes, with less variability in children's post-test scores. The study thus provides further support for the use of more cost-effective parental involvement approaches. The effects of the intervention also generalised to the children's language comprehension scores on the RDLS. An ANCOVA (with pre-test scores as covariate) was carried out in view of pre-test score differences in RDLS Expressive Language Scores. However, analyses were carried out on the raw scores of the RDLS and RAPT, thus losing the protection against the effects of maturation that standardised scores afford, and losing information about the normalisation of scores in the experimental and control groups. No information is provided regarding the on reliability of coding language sample measures.

* *E = experimental; C = control*

continued

TABLE 32 contd *Intervention studies – quasi-experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
McDade and McCartan, 1996	n = 18 (9E, age range 22–31 months, mean 27 months, SD 3.26 months and 9C, age range 21–30 months, mean 27 months, SD 3.24 months). A further two children allocated to the experimental group did not take part as there were no controls for them. Participants were selected from a larger group referred by health visitors because of concern about language delay. The criteria for inclusion in the study were normal general development and fewer than 50 words and no word combinations at 24 months, with no receptive or pragmatic delay. Seven subjects were excluded due to verbal agnosia/ autism. The participants all had normal hearing and no history of conductive hearing loss and met the criteria for specific expressive language delay. Eleven of the mothers had secondary schooling only. Seventy-four per cent of the sample were boys. All participants were Caucasian and English was the language of the home.	Quasi-experimental matched-pairs design. Subjects were selected and were then allocated to groups on the basis of availability. Experimental and control subjects were matched for age, gender, maternal educational level and maternal age.	Expressive language (vocabulary and syntax), receptive language and parent–child interaction	The study was designed to examine the effectiveness of training parents with the Hanen Program. The 12-week programme consisted of nine evening sessions (each lasting 150 min and designed to teach parents strategies using lectures, role play, focused discussions and videotapes) and three home visits. During the home visits parent–child interactions in free play were videotaped and viewed to provide parents with individual feedback.	The outcome measures were pre- and post-intervention criterion-referenced measures of maternal-child interaction, the number of information-carrying words used by the child and expressive language scores on the PLS-3 (Zimmerman *et al*, 1992). The results revealed that the experimental group children on average interacted and communicated more effectively with their mothers than the control group children. The experimental group also made significant and more marked gains than the control group on expressive language scores on the PLS-3 ($p < 0.01$) and in the total number of information-carrying words from the language sample transcripts. Post-intervention ratings indicated that the programme is highly acceptable to parents. **Comments** The results provide further support for the effectiveness of the Hanen Program for Parents. The scores on the PLS-3 reveal high levels of normalisation of the experimental group's expressive language scores as a result of treatment. However, there is no direct comparison between the progress made by the experimental group and that made by the control group, which would provide a stronger basis for evaluating the relative outcomes. The authors' checks on reliability of coding yielded coefficients in the range 0.71–0.89%.
Shelton, Johnson, Ruscello and Arndt, 1978 (Study 2)	n = 14 (4E1; 5E2; 5C) Participants were pre-school children with normal hearing who scored below average for their age on the Templin-Darley Articulation Test (Templin and Darley, 1969) and who had taken part in the authors' Study 1. The age range and mean age for each group at the time of the study are not given (though the study was carried out some 3 months (range 0–7 months) after Study 1) and no information is provided about gender balance.	Quasi-experimental design. Subjects were assigned to either a treatment group (parent-administered listening training directed towards correct identification of target sound or parent-administered listening training based upon reading and talking) or to a non-intervention control group.	Auditory listening skills; articulation and receptive language (syntax)	Children were given ten 15-min lessons administered by parents to provide additional training in the production of target sounds (in isolation, elicited by picture and auditory stimuli, by sentence completion and by conversation). The children moved from one level to the next upon achieving a criterion of 8/10 twice in succession.	Outcomes for one norm-referenced test (the Deep Screening Test of Articulation (McDonald, 1964)) and two criterion-referenced measures of sound production (one for sounds taught and the other for sounds not taught) revealed no difference between the listening training or the reading-talking training in terms of the children's response to the sound production training, though the gains for the sounds taught were larger than those for the sounds which were not taught. **Comments** The results reveal no significant differences between the two listening training approaches but the sample is very small, thus reducing statistical power (Cohen, 1992). In addition, the two groups in this study were not matched and there was no direct observation of parents and children at home to ensure treatment fidelity. Analyses were carried out on post-test–pre-test gain scores and on raw scores for standardised tests.

* *E = experimental; C = control*

continued

TABLE 32 contd *Intervention studies – quasi-experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Stevenson, Bax and Stevenson, 1982	n = 22 (12E, mean age 35 months, SD 4.5, and 10C, mean age 36 months, SD 4.1). The participants in this study were selected on the basis of the following criteria: age between 30–42 months, expressive language delay with a standard score of ≤ –1.5 on the Expressive Scale of the RDLS (Reynell, 1969), normal hearing, and fathers in manual social class occupations. The children all came from the inner city, 55% were boys, 23% came from immigrant families. The children had a range of problems in comprehension, expressive language and articulation/phonology and there were considerable within-group differences in level of delay and aetiology.	Quasi-experimental matched-pairs design. Subjects were assigned to either the experimental group (receiving home-based speech therapy) or non-intervention control group, with matching for day care arrangements, immigrant families, gender and non-verbal IQ.	Receptive language (vocabulary) and expressive language (syntax)	The study was designed to evaluate the effectiveness of home-based but therapist-administered speech therapy for young children with expressive language delay in an inner city area. Children in the experimental group received an individual treatment plan covering areas such as establishing a channel of communication, comprehension, expressive language, phonology and phonetics (for one child with dysarthria). They were seen in their homes by a speech therapist 22 times over a 6-month interval. Control group children did not receive any systematic treatment but their parents were given general advice at the initial assessment visit.	The outcome measures were the RDLS Expressive Scale, the EPVT (Brimer and Dunn, 1962) and the Griffiths Mental Development Scales for Hand/Eye Coordination and Performance (Griffiths, 1970). The results revealed that the children in both the experimental and the control groups made significant gains on the RDLS Expressive Scale ($p < 0.01$ and $p < 0.05$ respectively). However, there were no significant gains on either the EPVT or the Griffiths Scale. A direct comparison of the gains made by the experimental and control groups on the Expressive Scale using an ANCOVA with pre-test scores as a covariate failed to reach significance. However, the children in the experimental group made more consistent gains and none regressed. In contrast, four of the children in the control group regressed. **Comments** The study highlights the effects of maturation upon expressive language development, with the children in the control group making gains of the same order as the children in the treatment group receiving weekly home-based therapy. However, the pre-test scores for both groups on the RDLS revealed very marked levels of delay (of more than 2 SDs below the mean) which would be likely to increase the contribution of measurement error (in the form of regression to the mean) to the gains observed. However, even after intervention, the children in the experimental group continued to have significant and persistent expressive delay. The use of a larger sample to increase the statistical power of the comparison together with: (a) additional measures more sensitive to the goals of therapy, and (b) additional information about the presenting language problems would have been helpful.
Ward, 1994	n = 122 (49E1, age range 8–14 months, mean 10 months, SD 1.5; 9E2, age range 9–14 months, mean 11 months, SD 1.4; 9E3, age range 10–13 months, mean 12 months, SD 1.5; 52C1, age range 8–21 months, mean 11 months, SD 2.3; 8C2, age range 9–13 months, mean 11 months, SD 1.8; and 2C3, age range 9–13 months, mean 11 months). Participants all failed an early language screening test and were assigned to one of three groups. Children in Group 1 had receptive and expressive difficulties with associated listening difficulties, while those in Group 2 had receptive and expressive difficulties but no listening difficulties. Group 3 children had expressive difficulties only. An attrition rate of 17% is reported between the pre-intervention and the first follow-up assessment resulting in 43E1, 9E2, 2E3, 41C1, 4C2 and 2C3. Data from Groups 1 and 2 are reported here. No information about ethnicity, gender balance or SES is reported, but the study was carried out in an inner city area.	Quasi-experimental matched-pairs design. Subjects were selected following a post-screen assessment and were then allocated to one of the experimental groups or to one of the three control groups, matched for age, gender, locality, developmental and language quotients. The groups were also stratified on the basis of developmental and language quotients (high, medium and low scores).	Receptive language and expressive language	The study was designed to evaluate the effectiveness of the Ward Infant Language Screening Test, Assessment, Acceleration and Remediation (WILSTAAR) early intervention programme for young children (aged 10 months) with language difficulties. A speech and language therapist visited the homes of the children in the experimental groups four times in a period of around 4 months for 35 min and introduced a home-based programme which was individualised to meet the differing needs of those in Groups 1–3. Treatment encouraged the experimental group to develop selective attention to verbal/vocal input as a precursor to language development. The report does not indicate whether the programme has to be administered by a speech and language therapist or whether it could be delivered by a health visitor. The outcomes of the programme were evaluated at the end of treatment and then 1 and 2 years later using the REEL scale for receptive (and expressive) language (Bzoch and League, 1971). Treatment continued until the child was functioning within the normal range on the REEL receptive or expressive language, which required on average 4 monthly visits (range 2–7).	The results revealed that the children in the experimental groups increased their mean receptive language skill standard scores on the REEL by over 30 points on average after the 4-month programme. Follow-up assessments 1 year after the end of progress revealed that the scores of all of the treated children were within the normal range (i.e. standard score ≥ 85) while the average score for the control children continued to fall outwith normal limits. Two years after the end of intervention, at 3 years of age, all of the experimental group had age-appropriate play skills and all but two had continued to make progress and were functioning within the normal range in receptive language. (Both of the children who failed to make progress appeared to have emotional problems.) In contrast, 30% of the control group had been referred for speech therapy and their average scores continued to lie outwith the normal range, though there was considerable individual variation. **Comments** Although the sample size at follow-up was adversely affected by attrition (17% in the first year and a further 16% in the second), this study with its long-term follow-ups provides strong evidence for the effectiveness of a cost-effective early intervention programme which requires only four home visits. In view of the age of the children when they are treated (≤ 12 months) the programme has to be linked to a screening procedure. Replication with random allocation of children to treatment and control groups, a follow-up taking the children past school entry and investigation of the effects of the age of entry to the programme would be helpful.

* *E = experimental; C = control*

continued

TABLE 32 contd *Intervention studies – quasi-experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Warrick, Rubin and Rowe-Welsh, 1993	n = 28 (14E and 14C). (The study also included a further comparison group of 14 normally-functioning children.) The participants were selected from a group of kindergarten children who failed the Kindergarten Language Screening Test (Gauthier and Madison, 1978). Twenty-four of the subjects scored ≤ 12th percentile on a test of expressive language (either the Structured Photographic Expressive Language Test-II or Test-P) (Werner and Kresnick 1974; 1983) and four scored ≤ 25th percentile. Seven also scored ≤ 10th percentile on a test of language comprehension (either the TACL-R [Carrow-Woolfolk, 1985] or the TOLD-II [Newcomer and Hammill, 1982]). The children all had normal non-verbal IQs and were all non-readers. They had normal hearing and vision and no physical or emotional disorders. They were monolingual English speakers and came from similar SES backgrounds. No details of gender balance are reported. The children were thus a heterogeneous group comprising those with specific expressive and others with specific receptive plus expressive difficulties. There were no pre-test differences between the two groups on tests of phoneme awareness.	Quasi-experimental design. Subjects with language delay were assigned to either an experimental group (training in phoneme awareness skills) or a non-intervention control group.	Auditory discrimination/ listening skills (phoneme awareness)	The study was designed to evaluate the effectiveness of a programme to teach language-delayed children phoneme analysis skills which are associated with success in learning to read and spell. The experimental group took part in a structured training programme and were seen in two groups of seven by the third author twice a week for 20-min sessions over a period of 8 weeks. Each session consisted of 5 min of word play involving new goals or reviewing previous goals, followed by 10 min on structured phoneme awareness skills (e.g. syllable awareness, initial phoneme segmentation, rhyming and phoneme segmentation). The final 5 min of each session was spent on reviewing targeted skills, etc.	Criterion-referenced phoneme awareness measures were obtained pre-and post-intervention for both groups. The results revealed that the children who received trainng increased their scores on five of the six measures used ($p < 0.05$) after treatment while in contrast, the control group showed no improvement. Direct comparison between the treatment and control groups using post-test scores revealed that the treatment group had significantly higher scores on two tasks, rhyme and manipulation. In addition, after training there were no differences between the intervention group and the normally-functioning control group. A follow-up study 1 year later indicated that the intervention group continued to outperform the language-delayed control group not only on phoneme awareness tasks but also on reading ($p < 0.03$), indicating generalisation of the effects of training to decoding written words. **Comments** The results indicate that language-delayed children can be successfully trained in phoneme awareness skills and that they can 'catch up' with normally-functioning children in reading as well as in phoneme analysis.

* *E = experimental; C = control*

continued

TABLE 32 contd *Intervention studies – quasi-experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Whitehurst, Fischel, Lonigan, Valdez-Menacha, Arnold and Smith, 1991	n = 94 (mean age 28 months, SD 3.4). However, there was considerable attrition over a 3-year follow-up interval, complicated by differential attrition rates across the intervention and control groups. Data are reported here for n = 62 (25E and 37C) for whom complete pre- and post-intervention scores are available prior to a majority of the control group children enrolling in speech therapy programmes (31% of the control group received therapy at 34 months but this figure rose to 65% at 44 months). The participants in the study were referred by physicians or recruited via announcements in the media. They had normal hearing, no physical handicaps, normal intelligence (i.e non-verbal IQ ≥ 85 on the LIPS) (Leiter, 1976), normal receptive language (standard scores of ≥ 85 on the PPVT-R (Dunn and Dunn, 1981) but severe delay in expressive language (scores ≤ 2.33 SD on the Expressive One-Word Picture Vocabulary Test (EOWPVT) (Gardner, 1981). Children with evidence of autism, pervasive developmental disorder or physical disability were excluded. No details of ethnicity, SES, or gender balance are reported.	Quasi-experimental design. Subjects were allocated to the intervention group (receiving a home-based programme) on a 'first come, first served' basis, with assignment to the control group after the treatment group had been filled.	Expressive language	The study was designed to evaluate the effectiveness of a parent-administered home-based intervention programme for children with specific expressive language delay. The treatment programme consisted of seven written assignments which were tailored to meet the individual needs of families. The assignments covered activities to expand the child's expressive language skills (e.g. forced choice activities to encourage a shift from gestural to vocal communication; use of /wh-/ questions; incidental teaching; requirement of two-word combinations; description tasks during story time; use of open-ended questions during story time). The activities were designed to take account of the child's interests and utilised naturally-occurring contingencies. Parents were seen by a clinician in an office setting for 30 min every fortnight to receive an assignment. The sessions consisted of a review of the outcomes from the previous assignment and discussion of the new assignment. Role play (with the trainer playing the part of the child) and corrective feedback were used.	The results of follow-up assessment with standardised tests when the children were aged 34 months and after intervention was completed revealed that a significantly higher proportion of the intervention group had normalised their scores on the EOWPVT (72% vs. 41% of the control group, $p < 0.02$) and achieved higher scores on the expressive sub-scale of the ITPA (Kirk and Kirk, 1968) ($p < 0.01$). The intervention group children also used a higher percentage of target words in parent–child interactions after treatment. **Comments** The results reveal the effectiveness of the home-based intervention programme for expressive language delay, though the programme appeared to have little effect upon the subsequent emergence of phonological problems. It is unclear how much time parents spent on the assignments. The overall treatment effects observed may be an underestimate as children who 'dropped out' from the intervention group tended to be higher functioning than those who left the control group. Interestingly, although around a third of the control group were receiving other speech and language services there were no significant differences at 34 months between the control children who received community-based services and those who did not. The community-based therapy thus did not appear to advantage the control group children receiving it. Differences did emerge later on, at 65 months, but in an unexpected direction: control children receiving therapy had poorer scores than those who did not. The authors note that the control group children who received therapy did not have the most severe problems. However, the results from the three-year follow-up (n = 27) revealed that the advantages of intervention tended to 'wash-out' over time, with no difference between the two groups at 65 months. Phonological problems in both groups also resolved by the time the children were 65 months.

* *E = experimental; C = control*

continued

TABLE 32 contd *Intervention studies – quasi-experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Wilcox and Leonard, 1978	n = 24 (12E, age range 45–98 months, mean 68 months, SD 16.66, and 12C, age range 44–88 months, mean 58 months, SD 11.92). Participants all had normal hearing and articulation had all been diagnosed as language-disordered. All scored < 10th percentile on at least one of the following norm-referenced tests of expressive language: the NSST (Lee, 1969), the Carrow Elicited Language Inventory (CELI) (Carrow, 1974) or the Triota Speech and Language Test (Irwin, 1972). Subjects' MLU from a spontaneous language sample fell within the range 3.5–5.1 words. None produced well-formed /wh/ questions requiring the auxiliary verbs **is** or **does**. None of the children had been diagnosed as mentally retarded or neurologically impaired. 67% of the sample were boys (58% in the experimental group and 75% in the delayed treatment group). No information about SES is presented.	Quasi-experimental time series design. Subjects with language delay were assigned to either an experimental group (training in correct production of /wh/ question forms) or to a delayed treatment control group who were trained after the experimental group had completed their training. All but three of the children were randomly assigned to the groups.	Expressive language (syntax)	The study was designed to evaluate the effectiveness of a programme to train language-disordered children to correctly produce /wh/ question forms by means of modelling with instructions and differential reinforcement. Subjects were assigned to one of six experimental conditions and were trained using one of three /wh/ forms (**where**, **what** and **who**) and one of two auxiliary forms (**is** and **does**). The correct forms were modelled by the adult and a continous reinforcement schedule was used to reward correct responses by the child during training. When the subject achieved 20 consecutive correct responses on the these forms, a post-test was presented to assess generalisation of training to untrained auxiliary and /wh/ forms. No information is reported regarding the frequency or length of the sessions or the duration of the programme.	Comparison between pre- and post-intervention scores for the trained and untrained forms and auxiliaries revealed that the experimental group achieved significantly higher scores than the delayed treatment control group ($p < 0.001$). The control group also significantly increased their scores after they received training ($p < 0.001$). There was also evidence of generalisation of training to untrained /wh/ forms (**where** and **what** were the most generalisable) but there was little evidence of generalisation to untrained auxiliaries. **Comments** The results reveal the effectiveness of the experimental procedure and also the generalisability of training on **where** and **what**. The use of a time series design (with a delayed treatment group who made no gains during a baseline period extending over two pre-tests but then improved their scores following training) provides control for the effects of maturation. The authors' checks on reliability of coding yielded agreement of 97.8% for test sessions.
Zwitman and Sonderman, 1979	n = 22 (11E, age range 40–52 months; 11C, age range 42–54 months). Participants all had normal hearing and had a non-verbal IQ equivalent within the normal range (≥ 85) on the Minnesota Preschool Performance Scale sub-tests and normal comprehension (on sections A and B of the Assessment of Children's Language Comprehension). No information about gender, ethnicity or SES is reported.	Quasi-experimental design. Subjects were matched on their pre-test scores on the Developmental Word Sequencing Program pre-test and assigned (non-randomly) to either the experimental group (receiving training on syntax) or a non-intervention control group.	Expressive language (syntax)	The study was designed to evaluate the effectiveness of the Developmental Word Sequencing Program, designed to teach language disordered children syntax by means of imitation, modelling and differential reinforcement. The children in the experimental group were seen once per week for 45 min over a period of 2–6 months (average 5 months), which did not exceed 25 sessions. In addition, the children's mothers provided 10–15 min daily practice. The programme consisted of 11 sections ranging from basic naming vocabulary to four-word combinations each broken down into seven steps, and used picture cards to illustrate words and combinations. A spontaneous language sample for each child was used to ensure that structures missing from the Developmental Word Sequencing Program pre-test were also absent from the child's conversational speech. The level of agreement between pre-test usage and conversational usage was 94%.	The results revealed that children in the experimental group made significantly greater gains than the control group ($p < 0.05$). **Comments** The results indicate that all of the children in the experimental group improved the structure of their expressive language after treatment, though there was considerable variation in rate of progress, which appeared to be related to the level of skill at entry to the programme. Only one of the children successfully completed the whole programme, however. The programme appeared to be less successful in teaching descriptives, and the manipulation of cards representing structures in itself did not facilitate generalised use of structures in conversation (additional activities involving picture description were required). The effects of maturation were also evident, with five of the children in the control group making significant gains. Considerable variability in parental cooperation was also reported.

* *E = experimental; C = control*

TABLE 33 *Intervention studies – single-subject experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Bedrosian and Willis, 1987	n = 1M (aged 60 months) with hearing and motor development reported to be within normal limits for his age. Informal Piagetian Cognitive Assessment (Gill, 1979) and two sub-tests from the Clinical Evaluation of Language Functions (Semel and Wiig, 1980) indicated average/above-average cognitive ability and language comprehension. However, his MLU and type token ratio (Templin, 1957) indicated delay in expressive language with functioning in the 28–45 months range. The child also had problems in initiation, particularly memory- and future-related topics.	Multiple baseline across behaviours.	Pragmatics	The aim of the study was to increase the child's use of memory- and future-related topic initiation to more age-appropriate levels. The child was seen individually by a clinician in a clinic setting twice a week for 30 min each session over a 6-month period. The programme consisted of instruction, modelling, use of indirect and direct requests to elicit information and feedback within a communicative contexts which involved various practical activities and projects (e.g. making decorations) followed by discussion of the events that took place and of the events that would take place after the session. The criterion for ceasing treatment was the initiation of three memory or future-related topics for three consecutive 5-min probe sessions. Treatment commenced with memory-related initiations and when the criterion was reached, future-related initiations were treated.	The outcome measures were the number of 'here and now', future-related and memory-related initiations in 5-min probe sessions at the beginning of each of the 30-min sessions. The child achieved the criterion for success in future- and memory-related initiations, but the number of memory-related initiations decreased markedly below criterion after the withdrawal of toys. The structure of the child's expressive language also showed improvement, but in the absence of controls it is unclear to what extent the progress was due to maturation. Informal reports of generalisation of initiations to other settings were provided by the parents and the child's teacher. **Comments** The rationale for the criteria for success was unclear. Toys were used in early sessions but were withdrawn mid-way through intervention which may have adversely affected the number of memory-related initiations. Two out of three baselines were unstable (i.e. outwith ± 10% of mean baseline score), including the untreated control process ('here and now' initiations). The author's checks on reliability of coding yielded levels of agreement between 90–100%.
Camarata, 1993	n = 2 (1M and 1F, aged 46 and 51 months, respectively). The subjects had normal hearing and no neurological history but had phonological disorders with below-average scores on the GFTA (Goldman and Fristoe, 1986) and on samples from spontaneous conversation. Both children had received speech and language therapy.	Multiple baseline across behaviours (for three target phonemes in the case of the male subject) and across subjects (for production of target phoneme /l/).	Phonology	This study was designed to examine whether naturalistic conversation training is an effective means of improving the sound production of children with speech sound difficulties. The children were seen individually by a speech-language pathologist in a clinic for 45 min twice per week (for 14 sessions in the case of the girl and 35 sessions in the case of the boy) to work on individual speech sound targets arising from the initial standardised assessment and language sample. The naturalistic conversation training intervention model used in this study entailed the immediate provision of a correct model to the child following his/her incorrect production.	The outcome measures were the percentage correct sound productions across targets. Data are presented for three targets in the case of the male subject but for only one for the female subject, who dropped out prematurely after 14 sessions. The results reveal that both subjects achieved 100% production for target sounds. These levels generalised to use at home at the end of treatment and were maintained at follow-up some 9 months later. A 30-min training session by the author was sufficient for clinicians to carry out the programme. **Comments** Naturalistic conversation training resulted in the acquisition and correct use of the target sounds and improved speech intelligibility in spontaneous conversations not only with the therapist but with the children's mothers and an unfamiliar clinician in the children's homes. All of the baselines were stable (i.e. within ± 10% of mean baseline score). The authors' checks on reliability of coding of data yielded agreement in the range 83–92%.

* *M = male; F = female*

continued

TABLE 33 contd *Intervention studies – single-subject experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Connell, 1986a	n = 6 (age range 32–38 months, mean 36 months). (Note: data from two subjects only could be included as the number of baseline data points for the remaining four failed to meet the study design criteria of > 2.) The children had normal hearing and were not mentally retarded but were diagnosed as language disordered, though none had had therapy. Their scores on the PPVT (Dunn, 1965) were within 6 months of the normal range for their age but they produced only one-word utterances and the PLS (Zimmerman, Steiner and Evatt, 1969) revealed delay in comprehension (range 7–12 months).	Multiple baseline across subjects.	Expressive language (semantics)	The aim of the study was to examine whether production training and comprehension training could help language-disordered children learn to express semantic role meaning correctly in two-word utterances. Only the data for production training met the inclusion criteria. The two included subjects were seen individually in a clinic setting by a clinician 3–4 times per week for 40-min sessions for between 16 and 28 sessions in total (11 and 17 treatment sessions, respectively). Production training programme consisted of two steps: training the children to imitate sentences describing target sentences (a) with and (b) without modelling. Thirteen sets of picture pairs were used in the study, three for training and ten to serve as untaught generalisation probes. The criterion for success was 90% accuracy in three consecutive sets of ten trials. Reinforcement using tokens on a FR1 schedule was also used.	The outcome measures in this study were the percentage correct production of correctly-ordered sentences corresponding to the trained and the untrained pictures in response to probes administered on average after every third session, and the correct use in conversation. The results revealed that the two children reached the criterion for success for the production of semantic roles. The correct use of the semantic roles also generalised to untrained materials and to use in spontaneous conversation. **Comments** The results indicate that production training is an effective means of teaching language-disordered children the relationship between semantic roles and word order. The authors' checks on reliability of coding yielded agreement in the range 86–92% for the complete sample of six children.
Connell, 1986b	n = 4 (2M and 2F, age range 40–50 months, mean 45 months, SD 5.23 months). The subjects had normal hearing and were not mentally retarded but were diagnosed as language disordered, though none had had therapy. Their scores on the DSS (Lee, 1974) and the PLS (Zimmerman, Steiner and Evatt, 1969) revealed delay in expressive language (range15–30 months) and in auditory comprehension (range 7–12 months).	Multiple baseline across subjects.	Expressive language (syntax)	The aim of the study was to explore whether the predictions of the functional theory of language learning that there are two stages in the acquisition of the subject function (Givon, 1979) hold for language-disordered children. The subjects were seen individually in a clinic setting by a clinician 3–4 times per week for 30-min sessions for between 30 and 85 sessions. The programme consisted of eight stages and trained the children to say sentences with the correct subject form in response to picture stimuli by means of modelling, imitation, elicitation and reinforcement. Ten sets of four pictures were used in the study, five each for training a subject property and five to serve as untaught generalisation probes. The criterion for success was 80% accuracy on 20 consecutive responses over two sessions.	The outcome measures in this study were the percentage occurrence of the correct subject forms in response to probes administered after each session and the correct use in conversation. The results revealed that all four children reached the criterion for success for the five subject properties taught. The correct use of subject forms also generalised to use in spontaneous conversation. **Comments** The programme successfully trained subjects and provided support for the two-stage model of acquisition. However, there was considerable variation in individual patterns of learning. In addition, baselines for data combined over the five subject properties for two of the children were unstable (i.e. outwith ± 10% of the mean baseline score), though the baseline for one of these children was latterly stable for a period of over a 100 days prior to the introduction of training. The authors' checks on reliability of coding yielded agreement in the range 87–97%.

* *M= male; F = female*

continued

TABLE 33 contd *Intervention studies – single-subject experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Culatta and Horn, 1982	n = 3 (1M and 2F, age range 54–80 months, mean 67 months). (Note: another male subject from the sample was excluded from consideration here as his age (112 months) was outwith our inclusion criteria). The children's scores on the Columbia Mental Maturity Scale were all within normal limits. Two of the children had receptive language scores on the TACL (Carrow, 1973) some 7–10 months below their age and all had marked deficits in expressive syntax.	Multiple baseline across behaviours.	Expressive language (syntax)	The study was designed to evaluate the effectiveness of a four-step programme designed to facilitate generalisation of training of grammatical rules to spontaneous production. The subjects were seen individually twice per week for 45 min in a clinical setting by a clinician for 19–27 sessions. The programme used the following evoking strategies to support correct production by the child: modelling, role play and recreation of naturalistic events (e.g. getting ready for school). Modelling of targets by the clinician was reduced and the communicative context increased in complexity as the children progressed through the programme. Two grammatical rules were identified as targets for each child. Training was initiated for one rule at a time. The criterion for completing each step of the programme was 90% accuracy in productions of target rules in the first ten obligatory contexts on two consecutive sessions. Once the four steps had been successfully completed for the first rule, training for the second target commenced.	The outcome measures were the percentage of spontaneous productions of trained vs. untrained targets. The children achieved the required 90% accuracy rates for the first rules to be trained while untrained rules did not improve. When training for these untrained rules commenced the percentage of spontaneous productions improved and the 90% accuracy criterion was reached by all three children. Maintenance probes confirmed that these gains persisted after the withdrawal of treatment. **Comments** Training generalised to spontaneous productions and this highlights the effectiveness of using meaningful communicative contexts. Seven of the nine of the baselines were unstable (i.e. outwith ± 10% of the mean baseline score) but six of the seven were descending baselines. The authors' check on reliability of coding yielded agreement rates between 95–100%.
Dollaghan and Keston, 1986	n = 2, aged 70 and 75 months. (Note: a further two children from the sample were excluded from consideration here as their ages, 92 and 98 months, respectively, were outwith our inclusion criteria.) The subjects attended a special school for pupils with normal cognitive ability with a handicapping condition. The children were native English speakers with normal hearing and non-verbal cognitive ability in the low-average/average range. Both had problems with both language comprehension (on the PPVT-R) (Dunn and Dunn, 1981) and expressive language (e.g. syntax, reduced length of utterances, etc.) indicating specific language difficulties. No information about gender balance was provided.	Multiple baseline across subjects.	Receptive language (comprehension monitoring)	The study was designed to investigate the effectiveness of an intervention programme designed to improve comprehension monitoring skills, the skills used by listeners to detect and respond to breakdowns in comprehension (Markman, 1981). The children were seen individually by a clinician three times per week for 20 min over 4–5 weeks for 10–11 sessions. The programme had four phases: first, training in behaviours associated with an active orientation to listening; then training in responding when the message is obscured (e.g. by excessive rate or insufficient loudness); next, how to respond when the message contains inexplicit or ambiguous information; and lastly, how to respond when the message is excessively complex. Discussion, practice in identifying adequate and inadequate messages, role play and referential communication games were all used, in conjunction with positive reinforcement. Comprehension monitoring probes (eight commands, two adequate and six inadequate) were presented in random order during each session. The criterion for success was 100%. A generalisation probe of ten different and more complex messages was also presented on three occasions: before the first baseline session, at the 10th session and 3–6 weeks following the completion of treatment.	The outcome measures were the percentage of inadequate messages queried verbally by the subjects in response to the daily probes and the generalisation probes. The results revealed that 100% of the problem messages were queried verbally following treatment for both subjects. These gains were maintained following the withdrawal of treatment and were still in evidence at follow-up, some 3–6 weeks later. The treatment also generalised to different and more complex messages, though the performance on the generalisation probes was not as high as on the daily probes, particularly in the case of Subject 2. **Comments** The results indicate that training of comprehension monitoring skills can improve performance. But as with most criterion-referenced measures, there is no indication of the extent to which children's skills show normalisation. In addition, variability in the levels of generalisation was observed. The baseline for Subject 2 was also unstable (i.e. outwith ± 10% of the mean baseline score). The authors' checks on reliability of coding yielded agreement of 100%.

* *M= male; F = female*

continued

TABLE 33 contd *Intervention studies – single-subject experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Ellis Weismer and Murray-Branch, 1989	n = 4 (3M and 1F, age range 65–83 months, mean 70 months) with normal vision and hearing and age-appropriate cognitive abilities. English was the sole language of the home. Children's scores (≤ 1 SD below the mean) on one or more of a variety of standardised tests of language (e.g. TOLD – Primary (Newcomer and Hammill, 1982); Miller-Yoder Language Comprehension Test (Miller and Yoder, 1984); PPVT-R (Dunn and Dunn, 1981); and the Systematic Analysis of Language Transcripts (Miller and Chapman, 1985)) revealed that three had expressive language delay and one had specific problems in expressive language, phonology and language comprehension. All children were receiving speech and language therapy (on average for 6 months).	Alternating treatments with counter-balancing of the initial order of treatments across children.	Expressive language (syntax)	The study compared the effectiveness of two intervention approaches for problems in expressive syntax: modelling (Leonard, 1975) and modelling plus evoked productions of the target that do not simply imitate what the clinician says (Culatta and Horn, 1982). The children were seen in a clinic setting by clinicians for 1–2 sessions each week of 20–25 min over a 7–9-week period (i.e. 14–18 sessions). Individual targets were identified for each child using conversational samples (e.g. the auxiliary verb **is**; regular and irregular past tenses) and the two treatment procedures were alternated for each child using a semi-random ordering such that one type of treatment was not used for more than four consecutive sessions and only one treatment method was used during a given session. Elicitation tasks utilised picture descriptions (for Subjects A, C and D) and guessing pictured actions (for Subject B). Generalisation probes were presented at the end of each treatment session to establish the percentage correct production of target forms.	The outcome measures for this study were the percentage of correct responses to untaught generalisation probes for each target form during the baseline and treatment sessions. The results revealed that both approaches were effective for 3/4 subjects: two achieved 90–100% accuracy and the third 75% across both treatments. However, Subject C (who had problems with language comprehension as well as expressive syntax and had had only one month's therapy prior to the start of the study) made noticeably poorer progress (around 24% accuracy for modelling plus evoked production and less for modelling). Modelling plus evoked production tended to be associated with more stable (i.e. less variable) learning curves. **Comments** Modelling and modelling plus evoked production approaches both resulted in increased correct production of target forms. However the baselines for 3/4 subjects were unstable (i.e. outwith ± 10% of mean baseline score). In addition, the use of the alternating treatments design is vulnerable to treatment interference effects and also does not offer control against the effects of maturation (in the absence of a staggered multiple baseline across patients). The authors' checks on reliability of coding of data yielded agreement in the range 92–100%.
Ellis Weismer, Murray-Branch and Miller, 1993	n = 3 (2M and 1F, age range 27–28 months, mean 27 months). A fourth child was invited to participate but did not take part. The subjects had normal hearing, general cognitive abilities and language comprehension (on the SICD (Hendrick *et al,* 1975) for their age and they came from monolingual English-speaking homes but they were identified as 'late talkers' with restrictive productive vocabularies more than 2 SDs below the average for typically-developing age-peers and below the 10th percentile on the McArthur CDIs (Fenson, Dale, Reznick, Thal, Bates, Hartung, Pethick and Reilly, 1991). The children's MLU were ≤ 1 SD below the mean for their age (Miller and Chapman, 1981) but none had severe phonological difficulties.	Alternating treatments with counter-balanced order of presentation.	Expressive language (productive vocabulary)	The study compared the effectiveness of two approaches for developing productive vocabulary: modelling (Leonard, 1975) and modelling plus evoked productions of the target that do not simply imitate what the clinician says (Culatta and Horn, 1982) under group and individual instruction. The children were seen in a clinic setting by graduate student clinicians supervised by the investigators for two sessions each week of 60 min (40–45 min of group instruction and 15–20 min of individual instruction) over a 3-month period (i.e. 24 sessions). Language targets were identified for each child using object and action labels from the Early Language Inventory (Bates *et al,* 1986) which they could understand but not produce. Target and control words were balanced as far as possible for equivalency across treatment conditions. The two treatment procedures were alternated for each child using a semi-random ordering such that one type of treatment was not used for more than three consecutive sessions and only one treatment method was used during a given session. Checks were also made on the fidelity of the two treatment approaches.	The outcome measures were the percentage of correct probe productions per session; the number of different words produced in response to the probes per session; and the number of target words acquired under each treatment approach. Use of control words was also measured. The results indicated that children learned more target words and controls but revealed considerable individual variation and were in contrast to the earlier findings of Ellis Weismer and Murray-Branch (1989). One of the subjects made most gains in response to modelling, one to modelling plus evoked production and neither approach was effective in the case of the third. These findings were consistent across group and individual instruction. Scores on the SICD and the Early Language Inventory 6 months after pre-test also revealed post-test gains in receptive language (8–12 months) and vocabulary. However, the non-participating child made similar gains. Maternal reports indicated generalised use of target words but not control words outwith the clinic setting. **Comments** The results reflect the importance of considering treatment × aptitude interactions amongst subjects, although the gains made by the non-participating child suggest that the effects could be due to maturation. The effects of absenteeism upon the poor performance of one child is also discussed. The alternating treatments design is vulnerable to treatment interference effects and also does not offer control against the effects of maturation (in the absence of a staggered multiple baseline across subjects). The baselines were stable (i.e. within ± 10% of mean baseline score). The authors' checks on reliability of coding of data (second-level comparisons in the case of SALT and PEPPER analyses) yielded agreement in the range 90–100%.

* *M= male; F = female*

continued

TABLE 33 contd *Intervention studies – single-subject experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Gierut, 1990	n = 3 (all males), age range 49–58 months, mean 54 months. The children all had normal hearing, normal oral speech and motor ability and non-verbal IQs within the average/above-average range. However, they had specific phonological disorders, with below average scores on the GFTA (Goldman and Fristoe, 1986). None of children had received speech and language therapy and English was the sole language of the home.	Alternating treatments, with staggered multiple base-line across subjects and counter-balanced training order.	Phonology	The study provided a comparison of the relative effectiveness of two different treatment approaches for phonological disorder: treatment of minimal oppositions (Ferrier and Davis, 1973; Weiner, 1981) vs. treatment of maximal oppositions (Gierut, 1989). The children were seen by a therapist in a clinic setting for three 60-min sessions per week for an unreported number of weeks (which appeared to be around 30) and trained on nonsense word stimuli. Each child had two sound pairs identified for treatment, specific for each type of opposition. The alternating treatments design with a staggered multiple baseline across patients (Thompson and McReynolds, 1986) entailed presentation of both oppositions during each session, with a counterbalanced order of presentation. There was no direct training and intervention was based upon elicitation (imitation and spontaneous production). Reinforcement procedures were used during both imitative and spontaneous production phases.	The outcome measures were the percentage accuracy of treated and comparison sounds. The author concludes that the treatment of maximal oppositions resulted in a greater improvement in correct production of treated sounds and in generalisation to untreated sounds. The use of sound pairs which involved multiple and major class distinctions appeared most effective. **Comments** There was a good deal of variability in the results across subjects. Interestingly, the basis for concluding the superiority of treatment of maximal oppositions seems to have been the generalisation of learning to untreated sounds, rather than to a difference in outcomes for the treated sounds. The possibility that differential learning was the result of differences in the developmental order of acquisition of the sounds identified for each child was considered. However, the use of counterbalanced presentation order and the staggered multiple baseline offer a measure of control against the effects of treatment interference to which the alternating treatments design is vulnerable. Five of the six baselines were stable (i.e. within ± 10% of mean baseline score). The author's check on the reliability of coding yielded agreement in the range 86–97%.
Gierut, 1992 (Study 1)	n = 4 (3M and 1F, age range 42–54 months, mean 46 months). The subjects all had normal hearing, normal oral speech and motor ability and non-verbal IQs within the average/above-average range. However, they had specific phonological disorders, with below-average scores on the GFTA (Goldman and Fristoe, 1986). None of children had received speech and language therapy and English was the sole language of the home.	Alternating treatments, with staggered multiple baseline across subjects.	Phonology	The study provides an evaluation of the effects of minimal pairs treatment for phonological disorder (see Gierut, 1990). Subjects were treated directly by a speech and language therapist in a clinic setting for 60-min sessions, three times for 16–20 sessions using non-sense words. Each child was presented with two different formats of minimal pair treatment. The alternating treatments design with a staggered multiple baseline across subjects (Thompson and McReynolds, 1986) entailed presentation of both formats during each session, with a counterbalanced order of presentation. There was no direct training and intervention was based upon elicitation (imitation and spontaneous production). The criterion for success was 90% accurate production over three consecutive sessions within one minimal pair format (or 12 consecutive sessions, if earlier).	The outcome measure was the percentage accuracy of taught sounds for each treatment. The author concludes that teaching minimal pairs with two new phonemes unknown to the child is as effective or more effective than teaching one new phoneme. **Comments** There was a good deal of variability in the results across children and Subjects 11 and 12 in particular showed less response to the two new phonemes condition than to the one phoneme condition, which raises questions about the generality of the findings. Five of the six baselines were unstable (i.e. outwith ± 10% of mean baseline score). In addition, the use of counter-balanced presentation order and the staggered multiple baseline offer a measure of control against the effects of treatment interference to which the alternating treatments design is vulnerable. The author's checks on the reliability of coding yielded agreement in the range 87–91%.

* *M= male; F = female*

continued

TABLE 33 contd *Intervention studies – single-subject experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Hargrove, Roetzel and Hoodin, 1989	n = 1M, aged 80 months, known to speech and language services since 42 months. History of heart murmur and repeated middle ear infections but non-linguistic development was essentially normal. Receptive language, pragmatic skills, semantic relations, fine- and gross-motor skills, hearing and social skills within normal limits (e.g. on TOLD – Primary (Newcomer and Hammill, 1982), PPVT-R (Dunn and Dunn, 1981); CELI (Carrow, 1974)). However, marked problems in intelligibility and prosody were evident in the child's spontaneous speech (e.g. excessive and equal stress, prolonged and complex pitch changes and excessive rate).	Multiple baseline across behaviours.	Phonology/ articulation (prosody)	This study was designed to examine the effectiveness of a behaviour-based treatment programme for training prosodic skills. The goal of the programme was to produce a falling terminal contour, only one stressed word and a specific syntactic structure in a controlled context using a contrastive stress task (Baltaxe, 1984). The child was required to produce a targeted response contradicting the clinician's questions and the stress on the contradicted item was noted. Positive verbal feedback and cues were used by the clinician to elicit the target response. The child was seen at home by a certified speech and language pathology graduate student twice a day over a 9-day period spread over 2 consecutive weeks. The length of sessions was not reported. Three contradiction types were used to facilitate questions in which the subject, verb or object was incorrect. A criterion of 78% correct on targeted contradiction type probes or the completion of seven consecutive training sessions was required to move from one contradiction type to the next. Contradiction types were presented in a random order.	The outcome measure was the child's level of accuracy performance with generalisation probes (i.e. untrained stimuli) The results indicated that over 18 sessions the child reached the criterion for success (78% accuracy) on two of the behaviours and achieved 67% accuracy on the third. However, although the level of performance for object contradictions continued to increase during the maintenance phase, following withdrawal of treatment the levels of performance for subject and verb contradictions declined. **Comments** The results indicate that training in prosody skills can improve performance but the decline during the maintenance phase suggests that the treatment phase might usefully be extended. The baselines for verb and object contradictions were also unstable (i.e. outwith ± 10% of mean baseline score). The authors' checks on reliability of coding yielded agreement in the range 76–100%, indicating considerable variability. A check on treatment fidelity revealed adherence to the procedures 98.1–100% of the time.
Hedge and Gierut, 1979	n = 1M, aged 57 months. The subject's receptive vocabulary score on the PPVT (Dunn and Dunn, 1981) was within normal limits but spontaneous speech samples revealed problems with articulation and expressive syntax (e.g. MLU 3.8 words). He had not received any clinical speech and language therapy prior to the study.	Multiple baseline across behaviours.	Expressive language (syntax)	The study was designed to investigate whether training on the correct production of pronouns (**she**, **he** and **him**) and the verb form **are** would generalise to untrained stimuli. The subject was seen individually in a clinic setting by a clinician four times per week for 30-min sessions for a total of 45 sessions. The programme utilised imitation, modelling, prompts and reinforcement to evoke production of targets. Twenty stimulus pictures and sentences were used to train each target behaviour and 10–16 probe sentences were used for each to determine the extent of generalisation to untrained stimuli. The targets were introduced sequentially (in the order noted above) and once the subject achieved ten consecutively correct spontaneous productions on the first trained item of a selected pair the second item of the pair was introduced. When spontaneous responses to both items in a given pair reached 100% for around ten trials, probe trials were introduced. The criterion for success for probe trials was 100%. A response rate of less than this resulted in additional training.	The outcome measures were the percentage of correct responses for each target to probes in baseline, training and generalisation trials. The results revealed that the child achieved the 100% accuracy criterion for success for all four behaviours indicating the effectiveness of the training programme. In addition, training on the uncontractible auxiliary not only improved performance but also generalised to the uncontractible copula, and *vice versa*. The effects of training also generalised to the untaught items. However, there were marked differences in the number of trials required to reach the criterion, with initial items requiring more trials than later items, which seem to have been learned more quickly. **Comments** The results indicate that training of these grammatical features by means of modelling, imitation and systematic reinforcement can be effective and can also generalise to stimuli of the same classes which have not been taught. The stability of baseline measures (which provide experimental control for the effects of maturation) strengthens the conclusion regarding the effectiveness of the training programme. The authors' checks on reliability of coding yielded agreement of 100%.

* *M= male; F = female*

continued

TABLE 33 contd *Intervention studies – single-subject experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Hegde, Noll and Pecora, 1979 (Study 2)	n = 1M, aged 48 months who had received 11 weeks of treatment for articulation (though not for language) problems prior to the study. The child had normal hearing and oral speech mechanism and a score within normal limits on the PPVT (Dunn and Dunn, 1981). However, scores on the NSST (Lee, 1971) and CELI (Carrow, 1974) and performance on a spontaneous language sample indicated specific expressive language delay.	Multiple baseline across behaviours.	Expressive language (syntax)	The study was designed to investigate whether training on the correct production of syntactical features (the contractible copula and the uncontractible auxiliary) would generalise to other forms (the contractible auxiliary and the uncontractible copula). Training was also provided on a further form, the possessive /s/. The subject was seen individually in a clinic setting by a clinician 4–5 times per week for 45-min sessions for a total of 65 sessions. The programme utilised imitation, modelling, prompts and reinforcement (using sweets) to evoke production of targets. Fifteen stimulus sentences were used for the five target forms. Training targets were introduced sequentially and once the child achieved 11 consecutively correct spontaneous productions on trained items probe trials took place until the criterion of 90% generalisation to untrained sentences was met.	The outcome measures were the percentage of correct responses for each target to probes in baseline, training and generalisation trials. The results revealed that the child achieved the 90% accuracy criterion for success for all features indicating the effectiveness of training. In addition, training on the uncontractible auxiliary not only improved performance but also generalised to the uncontractible copula, and *vice versa*. The effects of training also generalised to untaught items in all cases. **Comments** The results indicate that training of these grammatical features by means of modelling, imitation and systematic reinforcement can be effective and can also result in generalised responses. The stability of baseline measures (which provide experimental control for the effects of maturation) strengthens the conclusion regarding the effectiveness of the training programme. The authors' checks on reliability of coding yielded agreement in the range 92–100%.

* *M= male*

continued

TABLE 33 contd *Intervention studies – single-subject experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Hemmeter and Kaiser, 1994	n = 1F, aged 25 months. (Note: a further three children from the sample were excluded from consideration here because of secondary delay, which is outwith our inclusion criteria). The subject had normal hearing and was very imitative but her scores on the SICD (Hedrick, Prather and Tobin, 1975) revealed a delay of some 17 months in receptive language and 9 months in expressive language. The child also had reported behaviour problems.	Multiple probe and multiple baseline across intervention strategies.	Expressive language and receptive language	The aim of the study was to examine the outcomes of training parents to use enhanced milieu training with their children with language delays. Parent and child attended two sessions of 45 min per week over 44 sessions. The parent was trained individually in a project centre playroom on strategies for arranging the environment to facilitate communication followed by strategies for incidental teaching (i.e. feedback and modelling the child's targets) once the criterion for success for environmental strategies was reached. Language targets were identified from analysis of interaction between the child and a research assistant and parental report and consisted of the use of the structures **want + noun**, **more + noun** and verbs such as **blow**, **jump** and **go**. Training sessions consisted of providing the parent with new information or feedback from the previous session and specific instructions on what to work on with their child in the practice session for 15 min; a 15-min video-taped practice session involving the parent, child and trainer, who provided coaching, if required; and a final 15-min session providing feedback and suggestions for work at home. The criteria for success were 80% correct in two successive sessions for environmental arrangement and incidental teaching strategies, use of responsive feedback by the parent on 70% of all opportunities during two successive sessions, and modelling of the child's targets 15 times in two successive sessions. Generalisation of training was assessed by four sessions of interaction between the child and a researcher at the end of both the baseline and intervention phases, and by six sessions of parent-interaction at home.	The outcome measures were the percentage correct use of strategies by the parents, the frequency of spontaneous child utterances, the frequency of spontaneous target use and the total number of targets used (i.e. prompted plus spontaneous). The results revealed that the parent required on average 16–17 training sessions to achieve criterion level on use of the enhanced techniques. The child also showed substantial increases in the use of targets in the training setting which generalised to the sessions with the research assistant and to the home. Re-testing on the SCID revealed that the subject made gains of 16 months in receptive language scores and 8 months for expressive language, reducing the child's language delay to around 4 months. **Comments** The other children in the study with secondary delays made greater gains on the criterion-referenced measures largely because the child considered here used significantly more spontaneous speech before intervention. However, this child made sizeable gains on a norm-referenced test indicating that the treatment was effective in normalising her language delay. Three of the four parent and One of the three child baselines were unstable (i.e. outwith ± 10% of the mean baseline score). The authors' checks on the reliability of coding yielded agreement on mean reliabilities in the range 74–92%, lower that is usual for this type of study.

* *F = female*

continued

TABLE 33 contd *Intervention studies – single-subject experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Kaiser, Hester, Alpert and Whiteman, 1995a	n = 2 (1M, aged 37 months, and 1F, aged 43 months). (Note: a further child from the sample was excluded from consideration here as he had a secondary delay which is outwith our inclusion criteria). The subjects' scores on the SICD (Hedrick, Prather and Tobin, 1975) revealed that both had delays in both expressive language (range 9–13 months) and receptive language (13–15 months). The PPVT (Dunn and Dunn, 1981) and the MacArthur CDI (Fenson *et al*, 1991) were also administered but the results were not reported.	Multiple baseline across subjects (trainer–parent–child triads).	Expressive language (syntax)	The aim of the study was to examine the effectiveness of teaching inexperienced trainers to teach parents to implement milieu teaching approaches with their delayed children. Parent training sessions were 30–60 min in length and were held twice a week in a clinic playroom setting for 10–11 sessions. Parents were trained in the milieu procedures (the Model, Mand-Model, Time Delay and Incidental Teaching Procedures) and in strategies for applying the procedures. Trainers' own training and their performance with parents and children were monitored. The Trainer Behaviour Code (Alpert and Kaiser, 1990) and the Milieu Language Intervention Code, a continuous observation coding system (Alpert, Tiernan and Fischer, 1988), were used to gather information about interactions between trainers, parents and children, fidelity to the treatment programme, the children's prompted and spontaneous production of targets (sentence structures) and generalisation to the home setting. The criterion for success was 80% correct of Milieu approaches in teaching episodes on two successive days.	The outcome measures were information given to parents by trainers; the methods used by trainers and the feedback given to parents; the frequency and percentage correct use of the milieu procedures by parents; and the prompted and spontaneous production of target language responses by the children. The overall results reveal that the trainer's parent training skills improved following intervention and that the parents learned to use the milieu teaching approaches achieving the criterion of 80% on all procedures. The children also showed increases in the number of language targets (prompted and spontaneous) used in conversation with their parents in the clinic setting. However, only one of the children showed generalised use of targets to the home setting. **Comments** The intervention of the trainers led to the parents successfully implementing the milieu procedures, with resultant improvement in their children's performances. However, there was a good deal of variability in the children's performances and in generalisation outwith the clinic setting. The baselines for both children also showed instability (i.e. outwith ± 10% of the mean baseline score). The authors' checks on the reliability of coding yielded agreement in the range 70–95%.
Kaiser, Hemmeter, Ostrosky, Alpert and Hancock, 1995b	n = 2, aged 42 months and 46 months, respectively. (Note: a further three children from the sample were excluded from consideration here because of secondary delay, which is outwith our inclusion criteria). One of the subjects was female but no information about the gender of the second was reported. The subjects' scores on the SICD (Hedrick, Prather and Tobin, 1975) revealed that both had delays in expressive language of 12–14 months. One of the children was diagnosed with developmental apraxia, the other with language delay.	Multiple baseline across subjects.	Expressive language (syntax)	The aim of the study was to examine the effectiveness of following group training for parents in the use of milieu teaching approaches with their delayed children with intensive individual home feedback sessions. Parents were first trained in groups of 2–3 in a conference room in a project centre. These sessions lasted for 60-90 min over eight sessions and each group had its own trainer, a graduate student. Parents were trained in the milieu procedures (the Model, Mand-Model, Time Delay and Incidental Teaching Procedures) and in seven strategies for applying the procedures which involved environmental arrangements. During the group training phase, the trainer videotaped parent–child interaction in the home for 15 min once per week and also provided additional instructions, feedback and coaching for a further 30 min. Intensive individual training took place in the family homes twice each week for 7–18 sessions after the end of the group training. Trainers provided specific instructions, feedback and coaching to parents and videotaped parent–child interaction. Feedback was provided to parents in the form of videos and graphs. The criterion for success was 80% correct use of Milieu approaches in teaching episodes on two successive days.	Outcome measures were the frequency of use of the environmental arrangement strategies by parents; the number of sessions to criterion by parents on the milieu teaching techniques; the prompted and spontaneous production of language targets by the children at home; and the percentage of child responsiveness to parent teaching at home. Both parents increased their use of environmental arrangement strategies and their frequency of use of milieu teaching techniques during group training and again during individual training, though there was some degree of individual variation. However, the criterion was only achieved following intensive individual training. One child's use of targets, particularly spontaneous productions, improved somewhat during group instruction but more dramatically during individual training. However, the second child produced most targets during group training. (This child did not have any spontaneous productions as her target was answers to /wh/ questions.) Both of these children were highly responsive to their parents during the baseline phase so showed little increase in responsiveness during training. **Comments** Parents improved their use of milieu techniques as a result of the intensive individual training but only one of the two children with primary language delay showed any differential improvement in the frequency of target productions as a result of intensive training, while group training was effective in the case of the second. There was no counterbalancing in presentation of treatments hence no control of treatment interference effects. Baselines for both children were unstable (i.e. outwith ± 10% of the mean baseline score). The authors' checks on the reliability of coding yielded agreement in the range 71–100%.

* *M= male; F = female*

continued

TABLE 33 contd *Intervention studies – single-subject experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
McGregor, 1994	n = 2 (1M, aged 56 months, and 1F, aged 60 months). Participants had mild-to-moderate expressive language delay and disfluent speech, though their phonological skills were age-appropriate. They scored 1–2 SDs above the mean on tests of receptive and expressive vocabulary PPVT-R (Dunn and Dunn, 1981) and the EOWPVT-R (Gardner, 1990). Word finding difficulties were apparent on errors on the EOWPVT-R and in discourse settings. Phonological encoding deficits were also evident on three syllable stimuli on the Goldman-Fristoe-Woodcock Auditory Skills Test Battery (Goldman, Fristoe and Woodcock, 1974).	Multiple baseline designs across subjects and behaviours with counterbalanced sets of training items.	Expressive language (word finding)	This study was designed to investigate the effectiveness of phonologically-based treatment for word finding. The children were seen individually in a clinic setting for two sessions each week over 24–26 sessions. Targets which were not in the subjects' receptive vocabularies were identified. Two sets of materials comprising eight target training words (represented by line drawings), eight phonologically related and eight semantically related were used. Half of the words were monosyllabic and half had three syllables. The sets were balanced and both trained and untrained (generalisation) items were probed. The treatment consisted of activities designed to reduce word finding difficulties by the use of visual cues to elaborate the storage of phonological output information (e.g. initial sound and number of syllables of target words). Each baseline and treatment session commenced with confrontation naming of all the words in the set. The criterion for success (and hence for concluding treatment) was no errors on at least four consecutive sessions.	The outcome measures were the total number of errors (semantic, phonological or no response) made in naming probe items in baseline, treatment and maintenance phases. The results revealed that the phonologically-based treatment decreased the number of errors and improved word finding performance by reducing both phonological errors and semantic substitutions on target words. The gains made were maintained three weeks after the cessation of treatment. There was little generalisation to the phonologically related control items and none to the semantically related control items. **Comments** The results support the view that the word-finding problems presented by the children in this study may have a phonological basis. However, baselines for both children were unstable (i.e. outwith ± 10% of mean baseline score). The author's checks on reliability of coding yielded agreement in the range 96–99%.
Olswang, Bain, Dunn and Cooper, 1983	n = 3 (2M and 1F, age range 23–40 months). (Note: a further girl from the sample was excluded from consideration here due to secondary delay). Subjects were diagnosed with moderate-to-severe speech and language delay at a university speech, language and hearing clinic. The children's scores on the SICD (Hedrick, Prather and Tobin, 1975) revealed that their receptive language skills were more advanced than their expressive skills (which showed an average delay of 14 months). Hearing and motor skills were within the normal range but scores on the Uzgiris and Hunt (1975) Scales of Psychological Development indicated some degree of general cognitive delay (particularly in the case of Subject 1). None had received any previous treatment.	Alternating treatments with counterbalanced order of presentation.	Expressive language (vocabulary)	The study was designed to compare the effectiveness of two treatments for vocabulary (lexical) learning: object manipulation and picture identification. The subjects were seen individually in a clinic setting by a clinician (one per child) three times per week for 30 min each session over a 7–8-week period. Two sessions each week were used for treatment to teach new single-word vocabulary (ten target nouns and ten target verbs) and the third probed the acquisition of the vocabulary by means of elicited production activities. Five nouns and five verbs were taught in a structured six-step programme using object manipulation stimuli and the remaining targets were taught by means of a six-step programme using picture identification. Both treatments were administered in each session (15 min each) and the order of presentation of treatments was counterbalanced. The target words were selected on the basis of their functional value to the child and were randomly allocated to the two treatment conditions. A further 20 control words were also identified.	The outcome measures were the number of responses per min for each treatment condition; the number of correct responses for each condition; and the number of correct spontaneous productions of target and control words. The results overall revealed that treatment increased single-word learning compared with baselines and untreated control words. But there were marked individual differences in the most effective treatment: two of the children learned most effectively by means of object manipulation but there was no differential effect in the case of the third. **Comments** The authors carried out a linear regression analysis which revealed individual differences in the rate of learning. The importance of treatment × aptitude interactions and of responding to children's individual learning styles is noted. However, the use of multiple treatments in the design (with the resultant possibility of treatment interference effects) and the instability in the baselines (i.e. outwith ± 10% of mean baseline score) pose problems for the interpretation of the findings. In addition, children made marked gains in learning untaught control words, possibly as a result of maturation. The authors' checks on reliability of coding yielded agreement in the range 85–100%.

* *M= male; F = female*

continued

TABLE 33 contd *Intervention studies – single-subject experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Olswang and Bain, 1985	n = 3M, age range 48–57 months, with normal hearing and age-appropriate production and comprehension skills. However, scores on tests such as the Compton-Hutton Phonological Assessment (Compton and Hutton, 1978), the Fisher-Logemann Test of Articulation Competence (Fisher and Logemann, 1971) and on PCC scores (Shriberg and Kwiatkowski, 1982) indicated that all of the subjects had specific phonological delay.	Multiple baseline across subjects plus ABA(B) withdrawal.	Phonology	This study examined whether treatment for phonological delay must be maintained until a high percentage of spontaneous productions of target sounds is achieved or whether the same outcomes can result from an earlier withdrawal of treatment. Graduate-student clinicians saw the children individually in a clinic setting 2–3 times per week for 50 min per session over a 38-week period. The sessions consisted of at least one session of direct treatment of target phonemes each week and one which in part consisted of conversational activity and elicitation techniques to obtain measures for analysis. Programmes used stimuli to elicit the production of sounds in isolation, syllables, words, phrases and sentences. Treatment was withdrawn following a performance at any one of three criteria (30%, 75% and 100% success) but was re-introduced if correct productions decreased or remained stable for three sessions and then continued until the next criterion was reached. Treatment and withdrawal were alternated for each child until the target phoneme was correctly elicited 75–100% of the time over three weeks without treatment.	The outcome measures for this study were the children's correct productions of their target sounds in untrained single words and in connected speech. The results from two of the subjects revealed that direct treatment could be withdrawn without affecting the continuation of the process of acquiring the phoneme once the children had achieved < 65% correct production of untrained single words. However, for the third child a higher level of performance (> 75% correct) was needed to maintain progress. **Comments** The results indicate that extended treatment need not necessarily be the rule: the withdrawal of initial treatment is recommended when the child reaches a criterion of 40–75% correct on single words, with longer-term treatment (to a criterion of 75–100%) if monitoring reveals that progress levels off. The use of a withdrawal condition followed by reintroduction of treatment, as in the case of Subject 3, provides additional support for the conclusion that the observed effects are due to the treatment. However, three points affect the generality and interpretation of these interesting findings: the small sample size and the variability in scores; the fact that two of the subjects required broadly the same number of trials to reach the 40% criterion as to reach the higher criterion; and the marked instability (i.e. outwith ± 10% of mean baseline score) of Subject 3 baselines. No percentage agreement rates for reliability of coding are presented though the data obtained by two clinicians were reported to be 'almost identical'.
Olswang, Bain, Rosendahl, Oblak and Smith, 1986	n = 2 (1M and 1F, aged 34 and 28 months, respectively). Children were referred to a university speech, language and hearing clinic on account of concern about delayed language development. English was the sole language of the home and hearing, motor skills and oral mechanism were normal. Receptive language skills on the SICD (Hedrick, Prather and Tobin, 1975) were also normal for their age but their expressive language skills were markedly delayed (by 10–12 months) and intelligibility was also a problem for both.	Alternating treatments design with counter-balanced order of presentation of treatments.	Expressive language (vocabulary)	The study was designed to explore the relationship between treatment and generalisation and the utility of dynamic assessment approaches for children with specific expressive language delay. The children were seen individually by a clinician in a clinical setting for three 30-min sessions each week for 30–33 sessions. The treatment programme consisted of teaching single-word vocabulary under three conditions: model only; model plus obstacle (where child was required to retrieve an object out of reach) and model plus elicitation. Seven items were randomly allocated to each condition and a further seven items served as untreated control items. Two sessions each week were used for treatment and the third was used to gather information about generalisation by means of elicited production and conversational activities.	The outcome measures were the percentage correct production of target items produced during treatment under the three conditions; the percentage correct production of target words under the two generalisation conditions (i.e. elicited production and conversational activities); and the percentage of the control words correctly produced. The results revealed a good deal of variation across the two children, particularly with regard to generalisation of treatment to elicited production and conversational activities where Subject 1 made greater gains than Subject 2. Overall, the children produced more correct target items under treatment conditions and the model plus elicitation treatment seemed to be the most effective. The dynamic assessment approaches helped to predict which of the children would best respond to treatment. **Comments** The alternating treatments design is vulnerable to treatment interference effects and also does not offer control against the effects of maturation (in the absence of a staggered multiple baseline across subjects), thus further complicating the interpretation of the findings. In addition, Subject 1's baselines were unstable (i.e. outwith ± 10% of mean baseline score) thus adding further complications. The authors' checks on reliability of coding yielded agreement in the range 81–100%.

* *M= male; F = female*

continued

TABLE 33 contd *Intervention studies – single-subject experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Olswang and Coggins, 1984	n = 2 (1M and 1F, aged 31 and 36 months, respectively). Children were referred to a university speech, language and hearing clinic on account of concern about delayed language development. English was the sole language of the home and hearing, motor skills and oral mechanism were normal. Receptive language skills on the SICD (Hedrick, Prather and Tobin, 1975) and PPVT (Dunn and Dunn, 1981) were also normal for their age but their expressive language skills were markedly delayed (by some 13–16 months).	Multiple baseline across behaviours, with randomised order of treatments.	Expressive language (semantics)	The study was designed to evaluate the effectiveness of three different treatment strategies on the children's production of three target two-term target relationships: a child-centred approach with adult modelling using natural language and expansion of child's spontaneous imitations; a clinician-centred approach, with adult direction and modelling of targets; and a combination approach with adult modelling, expansion, and spontaneous imitation. The children were seen individually each by a clinician (one per child) in a clinical setting for three 30-min sessions each week for 60 sessions over a 20-week period. Two sessions were used for treatment and one for gathering generalisation data. Treatment began with one target and the others were systematically introduced one at a time. Teaching on one strategy continued until the child failed to reach the criterion of spontaneous productions of two different target relationships or 20% more spontaneous imitations each week at which point a new treatment was introduced.	Outcome measures were the percentage spontaneous imitations and the number of spontaneous productions of the target two-term semantic relations under the three treatment conditions. The results revealed marked differences between the children. Treatment appeared to be effective for relations involving location and possession for Subject 1, but only for those involving possession for Subject 2. These gains were maintained following withdrawal of treatment. However, marked effects of maturation were also evident (treatment for some relations resulted in no greater changes than those which occurred during the baseline phase). Overall, there were no systematic differences between the three treatments. **Comments** The authors' linear regression analysis revealed individual differences in the time periods during which most rapid learning took place. However, the use of multiple treatments in the design (with the resultant possibility of treatment interference effects) and the instability in the baselines (i.e. outwith ± 10% of mean baseline score) pose problems for the interpretation of the findings. In addition, data for the untrained control words are not reported. The authors' checks on reliability of coding yielded close agreement (no percentage agreement reported).
Powell and Elbert, 1984	n = 5 (4M and 1F, age range 52–67 months, mean 60 months. (Note: a further girl from the sample was excluded from consideration here due to secondary delay). Subjects had normal hearing, no oro-motor problems, mental ages on the PPVT (Dunn, 1959) within the normal range and came from monolingual English-speaking homes. Their scores on the Templin-Darley Tests of Articulation (Templin and Darley, 1969) revealed marked problems in articulation.	Multiple baseline across subjects. Subjects were matched in pairs on the basis of age and mental age from the PPVT and were randomly assigned so that one child in each pair received Treatment A and the other, Treatment B.	Phonology/ articulation	This study examines whether teaching consonant clusters which are acquired at a later stage in phonological development (fricative + liquid) (Treatment B) leads to greater generalisation than teaching clusters which are acquired at an earlier stage in phonological development (stop + liquid) (Treatment A). The children were seen individually in a clinic setting for 30-min sessions over a 9-month period, with treatment lasting for 2–4 months. The frequency and total number of sessions is not given. The treatment consisted of a contrast of minimal pairs procedure involving verbal models, picture material, imitation and reinforcement of correct production of the target sounds selected for each child. The criterion for success for each contrast was 18/20 correct for 2–3 consecutive sets of 20 trials.	The outcome measure for this study was the children's correct productions of their target sounds in untrained single probe words. There was marked inter-subject variability but the results indicated that both Treatments A and B led to generalisation to treated and untreated cluster categories. **Comments** The baselines extended for up to 6 months for some subjects and thus unsurprisingly lacked stability (i.e. outwith ± 10% of mean baseline score), in most cases due to the effects of maturation. The absence of information regarding the frequency and number of intervention sessions makes it difficult to judge how intensive the treatment was (probe data was collected upon completion of Phase 3 of the training: the extent of training required to complete Phases 1 and 2 is not reported). The authors' checks on reliability of coding of data yielded agreement in the range 74–100%.

* *M= male; F = female*

continued

TABLE 33 contd *Intervention studies – single-subject experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Powell, Elbert and Dinnsen, 1991	n = 6 (4M and 2F, age range 59–66 months). Children were referred for treatment of phonological problems and scored at or below the 5th percentile of the GFTA (Goldman and Fristoe, 1986). They had normal hearing and scores within the average range on the PPVT-R (Dunn and Dunn, 1981), the TOLD – Primary (Newcomer and Hammill, 1982) and on tests of general intellectual ability.	Multiple baseline across behaviours.	Phonology/ articulation	The study examined the relationship between the participants' stimulability skills (i.e. the difference between spontaneous and imitative error rates averaged across phonemes) (Sommers, 1983), choice of treatment targets and generalisation of correct production of sounds. The children were seen in a clinic setting by the senior author on average for three 30-min sessions per week. The treatment consisted of a contrast of minimal pairs procedure involving verbal models, picture material, imitation and reinforcement of correct production of the target sounds selected for each subject. The criterion for success for each target was 90% accuracy over three sets of 20 items.	The outcome measures for this study were the percentage of correct responses during each session, generalisation data which were elicited by experimental and treatment probes, and follow-up data collected after the withdrawal of treatment. The results revealed that the treatment was effective and generalised to untreated sounds. The number of treatment sessions required to reach the criterion for success varied from 4–31. Progress was also maintained after the withdrawal of treatment. However, there was a significant correlation between stimulability skills and generalisation: limited generalisation of non-stimulable sounds was observed, indicating that direct treatment of such sounds is indicated. **Comments** The baselines for the majority of sounds were stable, though 4/20 baselines were unstable (i.e. outwith ± 10% of mean baseline score). The study also included untreated control processes as a further means of experimental control. The authors' checks on reliability of coding of data yielded agreement in the range 76–99%.
Warren, McQuarter and Rogers-Warren, 1984	n = 3 (1M and 2F, age range 35–43 months, mean age 39 months). The first subject's score on the Houston Test for Language Development (Crabtree, 1963) indicated a 22-month delay in expressive language, confirmed by a MLU of 1.1 morphemes. The second subject's scores on the PPVT (Dunn, 1965) and the Houston Test indicated delays of 19–20 months in both expressive and receptive language, associated with an MLU of 2.0 morphemes. The third child had an 11-month delay on the PPVT and restricted expressive language skills associated with an MLU of 1.4 morphemes. All of the children had normal hearing and intelligible speech but displayed very low rates of productive verbal behaviour in their university pre-school for language-delayed children.	Multiple baseline across subjects.	Expressive language	The aim of the study was to evaluate the effects of an intervention programme utilising mands, models and consequent events upon the expressive language of language-delayed children. The programme was carried out by teachers in the pre-school setting and consisted of two 45-min play sessions a week carried out over a period of 33–39 weeks. The teachers were trained in the use of mands (instructions to verbalise or questions requiring more than a 'Yes' or 'No' answer), models (imitative prompts) and positive feedback to encourage the children to verbalise. During the first part of the intervention the teachers were required to use the mand-model procedure at least ten times in a 15-min observation period. During the second part of the intervention, once stable rates of verbal response had been achieved, the subjects were require to respond to a mand or a model with a two-word or longer utterance. A second free-play session (30 min) provided a means of determining the extent to which the children's generalised to a setting in which the teachers were given no instructions to use the mand/model procedures. The programme also included a maintenance phase during which mands and models were faded out.	The outcome measures for baseline, treatment and generalisation settings were the mean numbers of teachers' non-yes/no questions; the mean numbers of mand and models; the percentage of child verbalisations followed by a contingent event; the total child verbalisations, the percentage of speech occasions responded to (responsiveness); and the number of initiations. The results revealed that the teachers did increase their use of mands, models and contingent events during the treatment phase and that the increases in teacher verbalisations were maintained during the maintenance phase, though there was a good deal of variability. The children also increased their numbers of verbalisations, initiations and responsiveness during the course of the treatment phase. Their MLUs also showed slight improvement, indicating that they produced longer utterances. Gains also generalised to a free-play setting. Initiation rates increased during maintenance for two of the three subjects, though the rates decreased during maintenance in the case of the third. **Comments** The mand-model procedure increased the number of the children's initiations, highlighting the fact that this is an active rather than passive model of learning. There was a good deal of variability in the rates of learning and the baselines for all three subjects were unstable (i.e. outwith ± 10% of the mean baseline score). The authors' checks on the reliability of coding yielded agreement in the range 75–100%.

* *M= male; F = female*

continued

TABLE 33 contd *Intervention studies – single-subject experimental designs*

Study	Subject characteristics*	Design	Areas of intervention	Study characteristics	Outcomes
Williams, 1991	n = 9 (5M and 4F, age range 44–69 months, mean 57 months, SD 8.58 months). The children had normal hearing, no oro-motor problems, scores within the average range on the PPVT-R (Dunn and Dunn, 1981) and errors on the sounds /s/ and /r/ on the GFTA (Goldman and Fristoe, 1986) indicating a range of severity of phonological disorder across subjects from mild to severe.	Multiple baseline across behaviours, with counter-balanced training order.	Phonology	The study examined the relationship between the participant's productive phonological knowledge and generalisation of correct production of sounds. The children were seen in a clinic setting by graduate student clinicians. Treatment was carried out over 9–42 sessions. Each training session consisted of five training sets (100 trials in total). The treatment consisted of three phases: imitation of picture items following the clinician's model with a continuous reinforcement schedule; imitation, etc, with a variable reinforcement schedule; and the spontaneous production of the training items in response to pictures without the clinician's model. Accuracy of production of 90% across three sets of 20 responses was required to move on to the next phase. The time taken to complete each session and the frequency of sessions was not reported. A generalisation probe consisting of 40 items of /s/ clusters and a similar number of /r/ clusters was presented three times during baseline and after every third treatment session) using delayed imitation elicitation. Training for each cluster continued until 70% accuracy on the generalisation probe was achieved or the subject completed a total of 21 sessions.	The outcome measures used were the generalisation probes. Five of the nine subjects achieved a criterion level of 70% accuracy or better for a cluster following training but there was considerable variation in the extent of generalisation. **Comments** The author notes that the subjects' productive phonological knowledge (e.g. production of individual sounds that comprise a cluster and the percentage sequential consonant production) prior to treatment appeared to account for some of the individual variation in the degree of generalisation observed. She further notes that definitions of phonological knowledge are too restrictive because of the effects of partial knowledge and that more consideration should be given to perceptual and acoustic data. Baselines for two of the subjects were unstable (i.e. outwith ± 10% of mean baseline score). The author's checks on reliability of coding of data yielded agreement in the range 87–100%.
Young, 1987	n = 2F, aged 52 and 53 months, respectively. The children had normal hearing and age-appropriate language comprehension scores on the PLS (Zimmerman, Steiner and Pond, 1979). Both had diagnosed expressive language delay and were enrolled in speech therapy. They had moderate-severe articulatory delay and impaired intelligibility on the AAPS (Fudala, 1970) and the Phonological Process Analysis (Weiner, 1979).	Multiple baseline across behaviours with counter-balanced order of presentation of dependent variables, use of an untreated control process and a follow-up session 6 weeks after the cessation of treatment.	Phonology/ articulation	This study investigated the effectiveness of treatment of two 'simplification' processes (weak syllable reduction and consonant cluster reduction) which can underlie misarticulations. These processes involve the omission of certain sounds when they occur in unstressed or weak syllables or consonant clusters, despite the fact that they can be correctly pronounced in stressed syllables (Ingram, 1976). The children were seen individually in a clinic setting for two sessions per week (Subject 1 for 24 sessions and Subject 2 for 18 sessions). The length of the sessions was not reported. Backward chaining procedures were used, together with rebus visual stimuli. In addition to the two processes trained a third, deletion of final consonant, was used as an untrained control process, thus increasing the internal validity of the study. Treatment for both processes consisted of modelling, imitation and fading of visual cues leading to spontaneous production and was presented in four steps of increasing difficulty. When the four steps for the first process had been completed, the second was introduced. The order of presentation of processes was counterbalanced across the two subjects.	The outcome measures were the percentage accuracy of production of taught and untrained (generalisation) words. The results revealed that both participants achieved or exceeded the 80% accuracy criterion for the correct production of both taught and untaught words for both weak syllable reduction and consonant cluster reduction. The follow-up data also confirmed that this level of correct production was maintained 6 weeks after the end of treatment. In contrast, the correct responses for the untreated control process never exceeded the baseline level of 20%. **Comments** The training programme demonstrated specific effects in that it was successful in improving the spontaneous correct production of the treated processes while leaving the untreated control process largely unaffected. Traditional therapy approaches have often adopted a sound-by-sound approach but the results here provide support for the alternative approach of utilising broader-based treatment programmes to treat simplification processes. It should be noted, however, that 4/6 baselines (including both for the untreated control process) fell outwith criterion for stability adopted here (within ± 10% of mean baseline score). The authors' checks on reliability of coding of data yielded agreement in the range 89–100%.

* *M= male; F = female*

TABLE 34 *Studies of concurrent validity of screening tests*

Study	Screening procedure	Criteria for language delay	Population sampled		Validity			Ranking			Comments
		Reference test, cut-off	No. given full testing/ screened sample	Age range (months)	Sensitivity	Specificity	LR	Replicability (of 17)	Validity (of 16)	Total (of 33)	
AFASIC, 1991 (clinical population)	AFASIC checklists	Clinical judgement; not fully specified	66/66	Not specified; test for 48–60 months	0.7	0.91	7.78	8	3	11	
Black *et al*, 1988 (clinical population)	Early Language Milestone Scale	REEL; score below 85 Bayley Mental Development Index; score below 85	44/48	8–22	0.83	1.00	*	12	6	18	Mothers came from low SES groups; half the children at risk of otitis media. Validity over 4 months.
Blaxley *et al*, 1983 (clinical population)	Fluharty Preschool Language Screening Test a. DSS 10th b. DSS 25th	DSS (Lee, 1974) a. DSS 10th centile b. DSS 25th centile	90/90	48–83	 0.36 0.30	 0.95 1.00	 4.1 *	11	8	19	
Bliss and Allen, 1984	Screening Kit of Language Development Standard English users a. 30–36 months b. 37–42 months c. 43–48 months Black English users a. 30–36 months b. 37–42 months c. 43–48 months	SICD; score less than 12 months below CA Clinical judgement; not fully specified	602/602	30–48	 1.0 1.0 1.0 0.90 0.89 0.95	 0.98 0.91 0.94 0.87 0.86 0.78	 45.00 9.89 15.17 6.84 6.16 4.36	11	10	21	Validation data based by authors on 513 children on whom reliability of observations was reached. Reconstructed data from original table given here.
Borowitz and Glascoe, 1986 (clinical population)	DDST language sector	PLS less than 0.7 of CA; with clinical assessment	71/71	18–66	0.46	1.0	*	14	9	23	Sample were referred children at risk of developmental problems.
Chaffee *et al*, 1990 (clinical population)	Minnesota Child Development Inventory	RDLS; more than 1 SD below mean: expressive receptive	152	24–87	 0.88 0.76	 0.45 0.63	 1.60 2.09	13	8	21	
Chevrie-Muller *et al*, 1993	Nursery school teachers questionnaire	Thirteen tests of receptive, expressive language; short-term memory and cognitive – motor skills; falling below 1 SD on at least seven tests, **or** falling below 2 SD on four tests.	480/940	39–45	0.79	0.78	3.6	6	11	17	

** LR not well defined (zero denominator)*

continued

TABLE 34 contd *Studies of concurrent validity of screening tests*

Study	Screening procedure	Criteria for language delay	Population sampled		Validity			Ranking			Comments
		Reference test, cut-off	No. given full testing/ screened sample	Age range (months)	Sensitivity	Specificity	LR	Replicability (of 17)	Validity (of 16)	Total (of 33)	
Clark *et al*, 1995	Clinical Linguistic Auditory Milestone Scale	SICD; below quotient of 80 Receptive language: 14–24 months 25–36 months Expressive language: 14–24 months 25–36 months	99/99	14–36	 0.83 0.67 0.50 0.88	 0.92 0.89 0.91 0.98	 10.28 5.89 5.5 42.0	13	13	26	Authors also provide tables for boys and girls separately.
Coplan *et al*, 1982 (clinical population)	Early Language Milestone Scale	REEL; PPVT; clinical judgement; cut-offs not fully specified	119/119	5–36	0.97	0.93	13.1	9	7	16	
Culatta *et al*, 1983	Story retelling	Screening TACL; cut-off not specified	199/199	Kindergarten to grade 1 Total Kindergarten readiness grade 1	 0.90 1.00 1.00 0.83	 0.79 0.72 0.57 0.86	 4.35 0.10 2.30 5.96	6	11	17	Activity not designed for a paraprofessional but at 10 min this could constitute a screening test.
Dale and Henderson, 1987 (clinical population)	Test of Early Language Development	Use of at least two standardised language tests, with clinical judgement: a. –1.5 SD b. less than 10th centile	85/85	36–83	 0.64 0.81	 0.68 0.68	 1.98 2.51	4	7	11	Unclear how clinical judgement was used. Referred sample of developmentally delayed children.
Dixon *et al*, 1988 (mixed sample)	Hackney Early Language Screening Test	RDLS; Symbolic Play Test (Lowe and Costello); Speech/language therapist's judgement	40/40	Not stated	0.94	0.95	20.8	8	9	17	
Dodge, 1980 (data extracts)	DDST; language sub-test	Using tests from: SICD; TACL; PPVT Cut-off –1 SD	486/486	Not stated	0.97	0.98	54.9	6	14	20	Author refers to PLS as a screen even though it is not for paraprofessionals. Sample of normal children.
Feeney *et al*, 1996 (Head Start population)	DDST; language sub-test	Meeting State criteria for eligibility to therapy services; not fully specified	183/199	36–59	0.80	0.95	14.93	8	5	13	Sample was enrolled for a Head Start programme.
Glascoe, 1991	Parent Evaluation of Developmental Status	Battelle Developmental Inventory screening test; expressive language sub-test: more than 2 SD below mean; Articulation Screening Test: pass/fail as in manual; [for 22 children, some of following additional tests: AAPS; the TOLD; SICD]; a failing score on Battelle or Articulation Screening Test or the battery = delayed.	157/157	6–77	0.72	0.83	4.24	11	10	21	Sample was drawn from those seeking paediatric care. Validation against other screening instruments; random selection of 22 children had more in-depth diagnosis to check out the use of Battelle and Articulation screen as reference tests.

continued

TABLE 34 contd *Studies of concurrent validity of screening tests*

Study	Screening procedure	Criteria for language delay	Population sampled		Validity			Ranking			Comments
		Reference test, cut-off	No. given full testing/ screened sample	Age range (months)	Sensi-tivity	Speci-ficity	LR	Replic-ability (of 17)	Validity (of 16)	Total (of 33)	
Haber and Norris, 1983 (Head Start popu-lation)	The Texas Preschool Screening Inventory	Hannah Gardner test of verbal and non-verbal langauge functioning; cut-off not specified.	53/53	48–60	0.92	0.98	36.9	9	6	15	Head Start sample; sampling procedure not specified. Hannah Gardner Test could itself be called a screen. Other validation samples reported here include a normative sample, but speech/language data lacking.
Illerbrun *et al*, 1985 (data extracts)	Fluharty Preschool Speech and Language Screening Test	TOLD;TACL; CELI: less than 15th centile.	136/136	58–77	0.65	0.94	10.77	5	11	16	Unclear how the three diagnostic tests are combined. Three months between administration of the screening tests and that of the diagnostic battery.
Klee *et al*, 1996 and 1997	The Language Development Survey	Clinical judgement; MSEL (Infant); either clinical concern and 1 sub-test of MSEL at –1 SD or no clinical concern and three sub-tests at –1 SD.	64/306	Mean 25.7	0.91	0.87	6.88	9	12	21	Author considers case for mailing parents with a screening questionnaire. See also predictive data at 3 years. Uses Rescorla's Delay 3 criterion on LDS (see below).
Law, 1994 with Law, 1993 (mixed sample)	Hackney Early Language Screening Test	RDLS; –1.5 SD	189/1205	Mean 31	0.98	0.69	3.17	14	9	23	
Levett and Muir, 1983	Levett-Muir Language Screening Test Receptive language Expressive language Speech	RDLS; GFTA; Language Assessment, Remediation and Screening Procedure; Cut-offs not specified.	14/140	23–39	 1.0 1.0 1.0	 1.0 1.0 1.0	 * * *	1	9	10	
McGinty, 1996 (mixed sample)	Mayo Early Language Screening Test	Language: RDLS; expressive and receptive combined score: –1 SD on either expressive or receptive section: 1;06–2 years 2–2;06 years 2;06–3 years 3–4 years 4–5 years Speech: EAT 3–5 years	200/200	18–60	 1.00 0.88 1.00 0.85 0.58 1.00	 0.55 0.80 0.78 0.69 0.65 0.62	 2.25 4.41 4.48 2.70 1.65 2.60	8	10	18	Sample made up of 120 referred children and 80 children not referred for therapy. Screen given by Public Health Nurse; diagnostic test by speech and language therapist, without knowledge of screen results.

** LR not well defined (zero denominator)*

continued

TABLE 34 contd *Studies of concurrent validity of screening tests*

Study	Screening procedure	Criteria for language delay	Population sampled		Validity			Ranking			Comments
		Reference test, cut-off	No. given full testing/ screened sample	Age range (months)	Sensitivity	Specificity	LR	Replicability (of 17)	Validity (of 16)	Total (of 33)	
Rescorla, 1989 (mixed sample)	The Language Development Survey	RDLS (expressive); 6 months delay	81/81	24–30				14	11	25	Mixed sample of clinical and non-clinical cases.
	Delay 1: less than 30 words **and** no combinations				0.50	0.97	17.60				
	Delay 2: less than 30 words **or** no combinations				0.76	0.89	6.90				
	Delay 3: less than 50 words **or** no combinations				0.89	0.86	6.35				
Rescorla, 1993	The Language Development Survey	Criterion 1: Bayley objects; none of 5 named at each of	92/92	Mean 24.6				9	10	19	High SES of sample. Study 2 of the cited paper quoted here as it has sufficient data.
	Delay 1				0.78	1.00	*				
	Delay 2				1.00	0.93	14.30				
	Delay 3				1.00	0.90	10.00				
		Criterion 2: Stanford Binet Intelligence vocabulary sub-test; none of 14 pictures named at each of									
	Delay 1				0.67	1.00	*				
	Delay 2				0.89	0.93	12.70				
	Delay 3				1.00	0.91	11.10				
Rigby and Chesham, 1981	A trial speech screening test for school entrants	Clinical judgement based on the Renfrew, RDLS and EAT; decision process not specified	438/438	54–57	0.8	0.93	12.09	8	13	21	Author explores effect of setting different cut-offs on the screen.
Scherer and D'Antonio, 1995 (mixed normal/ clinical population)	McArthur CDI: Toddler	Clinical judgement based on the PLS and Rosetti Infant-Toddler Language Scale; decision process not fully specified	60/60	16–30	0.76	0.91	8.53	11	9	20	Sample were 50% cleft and 50% non-cleft children.
Stevenson and Richman, 1976 with Richman *et al,* 1982	Screen used in prevalence study	RDLS, expressive language only; a. cut-off: score less than two-thirds of CA: 'severe' expressive delay	205/705	3 years	0.94	0.98	59.06	10	14	24	Authors conclude that screening for both language and behaviour difficulties would identify all language cases.
		b. cut-off: score less than age 30 months (i.e. 6-months' delay)			0.82	1.0	*				
Sturner *et al,* 1993a	Fluharty Preschool Speech and Language Screening Test	Study 1 1. TOLD – Primary; –1.5 SD 2. AAPS-R; 'severe' score	279/378	53–68				11	13	24	Borderline screen passes oversampled in criterion sample; high attrition before diagnostic testing.
		a. speech/language			0.43	0.82	2.4				Incomplete data.
		b. speech			0.74	0.96	18.5				
		c. language			0.38	0.85	2.53				

** LR not well defined (zero denominator)*

continued

TABLE 34 contd *Studies of concurrent validity of screening tests*

Study	Screening procedure	Criteria for language delay	Population sampled		Validity			Ranking			Comments
		Reference test, cut-off	No. given full testing/ screened sample	Age range (months)	Sensitivity	Specificity	LR	Replicability (of 17)	Validity (of 16)	Total (of 33)	
Sturner *et al*, 1993a	Fluharty Preschool Speech and Language Screening Test	Study 2 1. TACL; total score less than 10th centile 2. Templin-Darley Test of Articulation; –1 SD a. speech/language b. speech c. language	421/533	56–69	 0.31 0.43 0.17	 0.93 0.93 0.97	 4.40 6.14 5.67	11	13	24	Incomplete data.
Sturner *et al*, 1993b	Sentence Repetition Screening Test	a. Language ITPA auditory reception and auditory association sub-scales; Bankson Language Screening Test; both at less than 30th centile b. Speech AAPS; less than 15th centile	78/382	63–96	0.76 0.74	0.92 0.92	9.41 9.69	15	14	29	Diagnostic tests given within 4 months of screening. Oversampled at risk children for criterion sample; then used proportional weighting to factor back to the whole sample.
Walker *et al*, 1989	Early Language Milestone Scale	SICD; cut-off not specified 0–24 months 25–36 months	77/657	0–36	 1.00 0.69	 0.75 0.67	 4.00 2.07	8	9	17	Speech language pathologists used as the screeners may have raised the failure rate.
Ward and Birkett, 1994 WILSTAAR manual (See also Ward, 1984 and Ward, 1992)	WILSTAAR	REEL; cut-off 83–89 according to child's age band.	346?/1066	7–23	0.92	0.92	11.81	10	11	21	Data given here from manual as most consistent source.
Westerlund, 1997 with Westerlund, unpublished	Uppsala general language screening	PPVT; cut-off not specified. Swedish test of conceptual comprehension; clinical judgement: decision process described by author (NB. lack of standardised measures in Swedish).	44/2359	36–37	1.0	0.99	1158.5	6	7	13	Non-Swedish children excluded. Only 44/65 screen fails given diagnostic testing, thus data potentially weakened.
Whitworth *et al*, 1993	Parent questionnaire	Clinical judgement rating (not specified) as: a. moderate/severe problem b. mild/moderate/ severe problem	1106/6030	≈ 5 years	 0.81 0.69	 0.61 0.72	 2.06 2.47	5	14	19	Discusses also use of a teacher checklist for case finding. No follow-up of screen passes for teacher checklist, which precludes comparison of performances.

TABLE 35 *Comparison studies of concurrent validity between screening tests*

Study	Screening procedure	Criteria for language delay	Population sampled		Validity				Ranking			Comments
		Reference test, cut-off	No. given full testing/ screened sample	Age range (months)	Sensitivity	Specificity	PPV	LR	Replicability (of 17)	Validity (of 16)	Total (of 33)	
Allen and Bliss, 1987	Fluharty PLS Test NSST	SICD; receptive language at least 12 months below CA, or expressive language at least 12 months below CA.	182/182	36–74	0.60 0.92	0.80 0.48	0.33 0.22	3.15 1.79	13	12	25	White middle-class sample from day-care centres.
German *et al*, 1982 (clinical population)	Revised DDST: Expressive – Conservative Liberal Receptive – Conservative Liberal Developmental Profile II Expressive – Conservative Liberal Receptive – Conservative Liberal	SICD; scored as in manual.	84/84	Mean 41.7	 0.92 0.96 0.95 0.98 0.92 0.98 0.93 1.00	 0.49 0.14 0.45 0.14 0.72 0.42 0.62 0.36	 0.71 0.61 0.63 0.53 0.81 0.69 0.71 0.61	 1.79 1.12 1.74 1.14 3.30 1.68 2.44 1.56	11	8	19	Different cut-offs explored for the screen.
Glascoe and Byrne, 1993 (mixed normal/ clinical population)	Battelle Development Inventory Screening Test DDST II	Fluharty Preschool Speech and Language Screening Test; 3 sub-tests failed **or** VABS communication quotient more than 1.5 SD below IQ.	89/89	7–70	0.78 0.73	0.70 0.76	0.40 0.43	2.63 3.02	11	9	20	Sample 'socially at risk'; from day-care centres. Study also administered Developmental Profile II, but data for speech and language items not given separately.
Stokes, 1996	Parent questionnaire a. with comprehension item b. without comprehension item Nurses developmental screen	RDLS; –2 SD or language sample analysis; at stage I/II/III of syntax and/or phonology development.	398/398	34–40	 0.78 0.78 0.77	 0.91 0.95 0.97	 0.56 0.72 0.78	 8.33 17.21 28.17	11	12	23	Author notes deletion of comprehension item on parent screen leads to stronger tool.
Sturner *et al*, 1996	Sentence Repetition Screening Test Speech and Language Screening Questionnaire	a. Language ITPA auditory reception and auditory association sub-scales; Bankson Language Screening test; both at less than 30th centile b. Speech AAPS; less than 15th centile Language Speech	76/343	54–66	0.62 0.57 0.59 0.68	0.91 0.95 0.43 0.89	0.44 0.75 0.12 0.66	6.9 11.4 1.04 6.18	13 9	13 13	26 22	Two months between screen and criterion testing. Reporting of screen cut-off and reliability for Sentence Repetition Screening Test but not Speech and Language Screening Questionnaire.

Appendix 6

Reasons for exclusion of studies

Summary of excluded prevalence papers

Reason(s) for exclusion*	No. of papers
Excluded for a single reason	
4. No information about the sample size seen for full diagnostic testing (includes case of no diagnostic stage)	6
5. Sample not taken from a general but a clinical population	4
6. No clear criteria given for defining speech and language delay	2
Excluded for more than one reason	
Reasons 4, 5 and 6	2
Reasons 4 and 6	17
Reasons 5 and 6	1
Total number of excluded prevalence papers	**32**

* *Numbers refer to the inclusion and exclusion criteria sheets as given in appendix 4.*

Summary of excluded natural history papers

Reason(s) for exclusion*	No. of papers
Excluded for a single reason	
5. Not a prospective study	0
6. Follow-up interval less than 6 months	1
7. Pre-test and post-test language measures not available	3
Excluded for more than one reason	
Reasons 5 and 7	1
Total number of excluded natural history papers	**5**

* *Numbers refer to the inclusion and exclusion criteria sheets as given in appendix 4.*

Summary of excluded intervention papers

Reason(s) for exclusion*	No. of papers
Excluded for a single reason	
5. No detail on number of participants in each group	0
6. No pre-test or post-test intervention measures	2
7. Does not fulfil experimental design criteria	74
8. Does not provide details of treatment	0
Excluded for more than one reason	
Reasons 5, 6 and 7	2
Reasons 6 and 7	2
Total number of excluded intervention papers	**80**

* *Numbers refer to the inclusion and exclusion criteria sheets as given in appendix 4.*

Summary of excluded screening papers

Reason(s) for exclusion*	No. of papers
Excluded for a single reason	
6. No information about the sample size (i.e. no validation of the screen)	19
7. Sample drawn from those with high-risk neonatal histories	3
8. No clear criteria given for defining speech and language delay based on cut-off scores on a gold-standard test or objectified clinical judgement	10
9. No information given which allows calculation of concurrent validity for the speech and language items of the screen	8
10. Concurrent validity not expressed by both sensitivity and specificity	13

continued

Summary of excluded screening papers *contd*

Reason(s) for exclusion*	No. of papers
Excluded for more than one reason	
Reasons 7, 8 and 9	1
Reasons 8, 9 and 10	51
Reasons 8 and 9	38
Reasons 9 and 10	9
Reasons 8 and 10	4
Total number of excluded screening papers	**156**

* *Numbers refer to the inclusion and exclusion criteria sheets as given in appendix 4.*

Appendix 7

Instructions for coders

Coding strategy

The data extraction forms are designed to minimise coding errors by preserving as much of the original information from studies as possible, thus reducing the number of judgments that the coder is required to make.

Where possible, non-numerical information is coded into a numerical format. Lists of fixed and well-defined categories are provided to facilitate this. Additional categories will be added as necessary during the process of data extraction and coding. The categories used have been selected on the basis of knowledge of the literature and of likely effect modifiers and are sensitive enough to provide a means of testing hypotheses.

Information required to estimate effect sizes (e.g. means, SDs, *t, f, r,* chi-square, *p* values, etc.) is transcribed by coders. The DSTAT program is used to carry out the calculations required to combine effect sizes across studies.

Coding categories used in the project

The coding categories used in each of the four domains of the project are listed below.

Evaluation of prevalence studies

Studies reviewed in this domain of the project deal with the prevalence of speech and language delay in the general population up to the age of 16 years (i.e. excluding studies that deal exclusively with 'clinic' or hospital or special school/unit populations). A brief summary of the coding categories used is presented below.

Study details

This contains details of the authors' names, title of the book/paper/chapter/conference paper, etc.; reference of publication, date of publication.

Criteria for language delay

This indicates the test or procedure used to identify language delay and the cut-off scores used as the criterion for speech or language delay in a given study (e.g. –2 SD, < 5th percentile, fewer than 30 words used, etc.). The following categories are used to identify the particular areas of language delay: prelinguistic skills (including turn-taking, babbling, eye contact, concept of object permanence, development of visual and auditory attention), speech (including articulation, phonology, motor programming/praxis), expressive language (morphology, syntax, vocabulary), receptive language (syntax, vocabulary, and understanding of concepts), general language (unspecified), pragmatics/social use of language, and parent-child interaction. These categories are not mutually exclusive and problems in more than one area of language (e.g. receptive/expressive delays) are also indicated.

Sample

Details of the age range (months), mean age (months), SD of age, and size of the sample are included here. In addition, information about the balance of gender, SES and ethnicity of the participants is expressed in percentage form. Information about geographical location is also recorded, together with details of the inclusion and exclusion criteria of subjects (which relates to the relevance of the study) and of the sampling procedures used by the study. The latter are given by codes: (3), probability sampling; (2), non-probability sampling; (1), selected sample; (0), not specified.

Prevalence

Prevalence findings for each of the areas of language investigated are reported in this section, together with confidence limits, where these are available or can be calculated.

Quality rating

An overall rating of the quality of the study based upon study reliability (i.e. factors relating to the replicability of the study) and study validity (i.e. factors relating to internal and external validity) is provided, together with additional comments.

Evaluation of natural history studies

Studies reviewed in this domain deal with the natural history of speech and language delay (i.e. the outcomes of such delay in the absence of speech and language intervention). Studies in this section include not only longitudinal prospective studies specifically designed to examine

natural history but also the outcomes from non-intervention control groups with a test–retest interval of 6 months or more. A brief summary of the coding categories used is presented below.

Study details
This contains basic details of the authors' names, title of the book/paper/chapter/conference paper, etc.; reference of publication, date of publication.

Subject characteristics
Details of the age range (years), mean age (months), SD of age, and size of the sample are included here. In addition, information about the balance of gender, SES and ethnicity of the participants is expressed in percentage form. Information about geographical location and about co-existing disabilities are also recorded, together with any information about aetiology, details of inclusion and exclusion criteria (which relate to the relevance of the study), and information about the recruitment of subjects.

Design
The following categories are used to identify the design used in the study: prospective cohort study or treatment control group.

Areas of investigation
The following categories are used to identify the particular areas of language investigated by the study: attention control/concentration/listening skills, other prelinguistic skills (including turn-taking, babbling, eye contact, concept of object permanence), speech (including articulation, phonology, motor programming/praxis), expressive language (morphology, syntax, vocabulary), receptive language (syntax, vocabulary, and understanding of concepts), general language (unspecified), pragmatics/social use of language, cognitive abilities (e.g. IQ, memory), and parent–child interaction. These categories are not mutually exclusive and problems in more than one area of language (e.g. receptive/expressive delays) are also indicated.

Study characteristics
Information about who carried out assessments, the frequency and duration of follow-up and any checks on the fidelity of outcomes (i.e. of non-intervention) are contained in this section.

Outcome findings
The following categories of outcome measures are used: expressive language, receptive language, functional communication, cognitive skills, social skills and academic skills. These categories are not mutually exclusive and combinations may be reported. Results may also include measures of degree of impairment, disability, handicap, distress/well-being.

The following categories are used to indicate the type of outcome measure used: norm-referenced, criterion-referenced or qualitative ratings.

Details of pre- and post-test scores (means and SDs) are recorded, together with the results of any statistical tests or other data (e.g. percentages, frequency counts) which can be used to permit calculation of effect sizes.

Quality rating
An overall rating of the quality of the study is provided, based upon study reliability (i.e. factors relating to the replicability of the study) and study validity (i.e. factors relating to internal and external validity), together with additional comments.

Evaluation of effectiveness of intervention
Studies reviewed in this domain deal with the effectiveness of intervention for speech and language delay in the 0–7 years age range. A brief summary of the coding categories used is presented below.

Study details
This contains basic details of the authors' names, title of the book/paper/chapter/conference paper, etc., reference of publication, date of publication.

Subject characteristics
Details of the age range (years), mean age (months), SD of age, and size of the sample are included here. In addition, information about the balance of gender, SES and ethnicity of the participants is expressed in percentage form. Information about geographical location and about co-existing disabilities is also recorded, together with any information about aetiology and details of the inclusion and exclusion criteria (which relate to the relevance of the study), of the sampling and recruitment procedures and of the methods for allocating subjects to groups.

Design
The following categories are used to record the study design:

either (experimental designs)
- experimental studies (pre-test–post-test) with randomised control groups (i.e. no treatment or other treatment, randomised on the basis of individuals, classes or groups)

- quasi-experimental studies (pre-test–post-test) with non-equivalent control group (i.e. no treatment, non-randomised or pseudo-randomised selection of subjects) including multiple time series (with non-equivalent control group)

or

- Single-subject designs
  (a) withdrawal and reversal designs
  (b) multiple baseline designs
  (c) alternating treatment designs.

Areas of intervention

The following categories are used to identify the particular areas of language investigated by the study: attention control/concentration/listening skills, other prelinguistic skills (including turn-taking, babbling, eye contact, concept of object permanence), speech (including articulation, phonology, motor programming/praxis), expressive language (morphology, syntax, vocabulary), receptive language (syntax, vocabulary, and understanding of concepts), general language (unspecified), pragmatics/social use of language, cognitive abilities (e.g. IQ, memory), and parent–child interaction. These categories are not mutually exclusive and problems in more than one area of language (e.g. receptive/expressive delays) are also indicated.

Study characteristics

Information about the setting for the intervention (e.g. clinic, hospital, laboratory, school, nursery, home), who carried out assessments, who carried out interventions, the frequency and duration of treatment sessions and fidelity to treatment programme is contained in this section.

Outcome findings

The following categories of outcome measures are used: expressive language, receptive language, functional communication, cognitive skills, social skills and academic skills. These categories are not mutually exclusive and combinations may be reported. Results may also include measures of degree of impairment, disability, handicap, distress/well-being.

The following categories are used to indicate the type of outcome measure used: norm-referenced, criterion-referenced or qualitative ratings.

Details of pre- and post-test scores (means and SDs) are recorded, together with the results of any statistical tests or other data (e.g. percentages, frequency counts) which can be used to permit calculation of effect sizes.

Quality rating

An overall rating of the quality of the study is provided, based upon study reliability (i.e. factors relating to the replicability of the study) and study validity (i.e. factors relating to internal and external validity), together with additional comments regarding information provided about cost-effectiveness and authors' comments about the components of treatment regimen considered to have most impact on intervention result.

Evaluation of the accuracy of screening procedures

Studies reviewed in this domain deal with screening for speech and language delay in the 0–7 years age range. A brief summary of the coding categories used is presented below.

Study details

This contains basic details of the authors' names, title of the book/paper/chapter/conference paper, etc., reference of publication, date of publication. It also indicates whether the study is part of a series of papers on a particular test.

Screening procedure

This section deals with:

- the areas of language covered by the screening test: prelinguistic skills (including turn-taking, babbling, eye contact, concept of object permanence, development of visual and auditory attention), speech (including articulation, phonology, motor planning/praxis), expressive language (morphology, syntax, vocabulary), receptive language (syntax, vocabulary, and understanding of concepts), pragmatics/social use of language and parent–child interaction.
- the name of the test used; recommended age-range; time taken for administration; source of information (given by: (1), child; (2), parent/carer; (3), doctor/paediatrician; (4), health visitor/teacher/nursery; (5), other – specified); the method used to obtain data (given by: (1), observation of behaviour; (2), direct assessment using a test; (3), parent/carer report; (4), other — specified); whether the test assesses language only (monophasic test) or covers other areas of development (given by: (1), monophasic; (2), multiphasic); details of any costs provided by the authors.

Criteria

This section deals with the criteria used to define language delay and identifies the reference or 'gold-standard' test used, the level of cut-off used on the screening test (e.g. –2 SD, < 5th percentile, fewer

than 30 words used), and any rationale provided by the authors for the use of such a cut-off.

Sample

Details of the age range (years), mean age (months), SD of age, and size of the sample are included here. In addition, information about the balance of gender, SES and ethnicity of the participants is expressed in percentage form. Information about geographical location is also recorded, together with details of the inclusion and exclusion criteria of subjects (which relate to the relevance of the study) and of the sampling and recruitment procedures. The latter are given by codes: (3), probability sampling; (2), non-probability sampling; (1), selected sample; (0), not specified.

Reliability

Details of inter-rater and test–retest reliability are contained in this section, marked as high: above or equal to 0.8; moderate: above or equal to 0.5; or low.

Validity

Information about concurrent validity is recorded by means of the 2×2 contingency table, allowing calculation of sensitivity, specificity, PPV, and LR. Other validity measures are noted if reported (e.g. construct validity, content validity).

Quality rating

An overall rating of the quality of the study is provided, based upon study reliability (i.e. factors relating to the replicability of the study) and study validity (i.e. factors relating to internal and external validity), together with additional comments.

Training for coders

To ensure that coders can use the forms, conventions and procedures of the synthesis consistently and reliably, the following iterative process is used.

(1) Coders are given a sample of five papers to read.
(2) Working in pairs in each of the two project sites, each item on the form and its description in the code book is read and discussed.
(3) A further sample of five studies is chosen to 'test' each of the four forms in each project site. Further additions, deletions or modifications to the data extraction proforma are discussed.
(4) A study is coded by everyone, with each coder recording how long it takes to code each item (this provides estimates of how long it takes to code individual items and complete studies).
(6) Coded forms are compared and the discrepancies are identified and resolved.
(7) Forms and code book are revised as necessary.
(8) Another study is coded and reviewed, and so on. Steps 4–8 are repeated until apparent consensus is achieved.

Reliability of coding

The percentage agreement rate for two independent coders was calculated for each of the four domains, i.e. prevalence, natural history, intervention and screening, and is shown below.

TABLE 36 *Mean percentage agreement rates across domains, based on point-by-point agreement between two independent coders*

Domain	% of studies used in reliability check	Overall % agreement rate (range per item)
Prevalence	19	84.8 (0–96)
Natural history	14	85.4 (50–100)
Intervention	9	89.0 (62.5–100)
Screening	13	90.2 (55–100)

Appendix 8

Data extraction forms

Contents

A) EVALUATION OF PREVALENCE STUDIES

(Notes for coders are in italics)

REVIEWER: (Initials)

DATE OF CODING:

A. STUDY DETAILS

TITLE:

AUTHOR(S):

SOURCE_NAME AND REF:

SOURCE_DATE:

PART OF A SERIES OF PAPERS ON THIS TEST (Y/N/NK) *(This is a check for multiple or related reporting about the same cohort)*

B. CRITERIA FOR LANGUAGE DELAY

Diagnostic test(s) used (e.g. at final stage)	**Areas of language probed by test**	**Cut-off (criteria for delay)**

C. SAMPLE

Sampling/recruitment procedures *(Use 0–3):*

Number in pre-screen sample having full diagnostic testing: and as % of screened sample:

How was sample chosen for full diagnostic testing? *(Use 1–4):*

Sample descriptions

	Original Sample	**Pre-screen sample**	**Diagnostic sample**
Nature of sample Normative/clinical/mixed			
Number			
Exclusions *(Note any reasons below)*			
SES *(Indicate here % balance of SES [non-manual/total] etc., e.g. 23% of sample non-manual)*			
Ethnicity *(Indicate here % balance ethnicity [ethnic groups/total] e.g. 17% of sample black)*			
Gender *(Indicate here % gender balance [males/total] e.g. 56% of sample male)*			
Age range *(Months)*			
Mean age *(Or median, if reported instead; months – round to nearest month)*			
SD age *(Months)*			

Geographical location *(Indicate here the country in which the study was carried out. Indicate below whether sample is from one town/area, etc.)*

(1) One county/health authority (i.e. more than one town/city)
(2) Sample from one town/city
(3) Other (specify)

Urban *(Indicate here % balance of urban (e.g. urban/total):*

No. of children included/excluded	**Reasons** *(free text, e.g. developmental delay, bilingual, etc.)*
Exclusions	
Inclusions	

Net no. of children excluded from sample *(and as % of original sample):*

D. PREVALENCE

Area of language	Prevalence – males	Prevalence – females	Total prevalence
Prelinguistic skills			
Speech			
Expressive language			
Receptive language			
Expressive + receptive language			
Pragmatics			
General language *(unspecified)*			
Total overall estimate of prevalence *(including any confidence limits provided by the study authors)*			

E. QUALITY RATINGS

Study reliability *(Factors relating to quality of reporting)*		**Study validity** *(Factors relating to internal and external validity)*	
	Score		**Score**
Subject		**Population**	
Details of age (report one of: mean, range or SD)	1	Sampling procedure	
Details of gender	1	0–3 as in form above:	
Report details of one of: ethnicity, SES	1	0	0
		1	1
Procedures		2	2
For each reference test:		3	5
Details of the cut-off	4		
		Sample size given diagnostic testing:	
For the use of clinical judgement:		50/50–150/150+	0–3
Method given for reaching decision	2		
		Reference testing	
		Choice of children to give full diagnostic test to:	
		Only pre-screen fails	1
		Selected pre-screen fails and some pre-screen passes (selection not specified or not stratified)	2
		Selected pre-screen fails and passes – specified as stratified or random	3–4*
		All of sample pre-screened	3–4*
		* Score 4 where attrition less than 20%	
Score out of 9:		Score out of 12:	
Grade low: **Grade medium:** **Grade high:**		**Score summary:** _____of 9 _____of 12 _____of 21	

B) EVALUATION OF LONGITUDINAL NATURAL HISTORY STUDIES

REVIEWER: DATE OF CODING:
(Initials)

A. STUDY DETAILS

TITLE:

AUTHOR(S):

SOURCE_NAME AND REF: SOURCE_DATE:

PART OF A SERIES OF PAPERS ON THIS TEST (Y/N/NK) *(This is a check for multiple or related reporting about the same cohort)*

B. SUBJECT CHARACTERISTICS

Age range *(Months):*

Mean *(Or median, if reported instead)* **age** *(Months – round to nearest month):*

SD *(Months):*

Size of sample:

Sampling/recruitment procedures

(3) Probability (or representative) random sampling *(i.e. were subjects recruited via sampling from the population using simple random sampling, systematic sampling, stratified random sampling, or cluster sampling? Note this procedure requires the researcher to have access to a list of the total population)*

(2) Non-probability (i.e. non-random) sampling *(e.g. quota sampling, convenience sampling)*

(1) Selected *(e.g. selected not sampled from group of already diagnosed children, or where professional nominates a proportion of children who attend a routine clinic without knowledge of case status)*

(0) Procedures are insufficiently recorded

% Take-up rate *(Indicates the % of children sampled who actually participate in the study):*

Inclusion/exclusion criteria *(Note information provided by the author(s) regarding criteria for speech and language delay used in the study):*

No. excluded from sample *(Indicate % excluded due to any of the following):*

(1) Developmental delay (2) ESL (3) Other (specify)

Co-existing disabilities *(Indicate any co-existing disabilities and the % balance: disabilities/total):*

(1) Neurodevelopmental immaturities (e.g. hand–eye coordination):

(2) Specific learning difficulties:

(3) Medical complications (e.g. conductive hearing impairment: specify):

(4) Behaviour:

(5) Others (specify)

Description of sample

SES *(Indicate here % balance of SES [non-manual/total] etc., e.g. 23% of sample non-manual):*

Ethnicity *(Indicate here % balance ethnicity [ethnic groups/total], e.g. 17% of sample black):*

Gender *(Indicate here % balance gender [males/total], e.g. 56% of sample male):*

Geographical location *(Indicate here the country in which the study was carried out.*

Indicate below whether sample is from one town/area, etc.)

(1) One county/health authority: (i.e. more than one town/city)

(2) Sample from one town/city

(3) Other (specify)

Urban *(Indicate here % balance of urban, e.g. urban/total):*

C. DESIGN *(Check that the study meets one of the three design criteria below)*

1. Prospective cohort study

2. No-treatment controls with test–re-test interval of 6 months or more

3. Study of predictive validity of screening test

D. AREAS OF INVESTIGATION

Indicate using the table below (e.g. ***attention control/concentration/listening skills, prelinguistic skills*** *[such as turn-taking, babbling, eye contact, concept of object permanence, development of visual and auditory attention, etc.];* ***speech*** *[including articulation, phonology, motor planning/praxis, etc.];* ***expressive language*** *[including morphology, syntax, vocabulary, other aspects of semantics, etc.];* ***receptive language*** *[including syntax, vocabulary, understanding of concepts, etc.]; general language [area unspecified];* ***pragmatics/social use of language, cognitive abilities*** *[e.g. IQ, memory], and* ***parent–child interaction****. Note that these areas are not mutually exclusive: indicate as many as are relevant).*

Area of language	**Investigated (Yes/No)**
Attention control/concentration/listening skills	
Prelinguistic skills	
Speech	
Expressive language	
Receptive language	
Expressive + receptive language	
Pragmatics	
General language (unspecified)	
Cognitive abilities (including IQ, memory, etc.)	
Parent–child interaction	
Other (Specify)	

E. STUDY CHARACTERISTICS *(Information here addresses issues relating to possible bias which may affect construct validity)*

Who carried out assessments? *(i.e. was it the researcher or therapist, or someone independent? Was the assessor 'blind' to the status of the children seen?):*

Duration of (test–re-test) follow-up interval	
Information about outcome fidelity *(i.e. was specific non-intervention maintained?)*	

Details of any treatment *(Free text):*

F. OUTCOME FINDINGS

Outcome measures *(Indicate the range of output measures used. More than one category may be used, if multiple measures were used):*

IQ Expressive language Receptive language Functional communication

Cognitive skills Social skills Academic skills

Area covered *(If WHO definitions are used):*

Impairment Disability Handicap Distress/well-being

Instrumentation *(e.g. type of language measures used):*

Norm-referenced Criterion-referenced Qualitative ratings

Case status at onset of study *(% of children with speech/language delay. Note separate % for males and females)*

Case status at end of study *(Note separate % for males and females)*

Measures used
(Complete table below. Include name of any test(s)/variables)

Test used or other dependent variable *(Specify)*	**Pre-test**		**Post-test**	
	Mean/median *(Indicate which)*	**SD**	**Mean/median** (Indicate which)	**SD**

Details of any statistical tests used *(Insert appropriate values from study, or use a separate sheet if preferred):*

t F Chi-square *p*

% Frequency counts Effect size (if reported)

G. QUALITY RATINGS

Study reliability *(Factors relating to replicability of the study: score information provided as indicated below)*	**Score**	**Study validity** *(Factors relating to internal and external validity: score information provided as indicated below)*	**Score**
Subject		**Sampling Procedures**	
Details of Age:		Random (i.e. population study)	3
Mean	1	Non-random	2
Range	1	Selected	1
SD	1	Not given	0
Details of SES	1		
Details of gender	1	**Attrition**	
Details of ethnicity	1	< 30%:	2
Details of basis for subject classification		>30%:	
(e.g. criteria for language-delay etc.)	1	Replaced or controlled	1
		Nothing done	0
Materials			
Description, list or names of materials used	1	**Inclusion/exclusion criteria**	
		Given	1
		Not given	0
Procedures			
Details of who measures outcomes	1	**Definition of subjects' speech and language problem**	
		Given	1
Instrumentation		Not given	0
Standardised (diagnostic or screening)	1		
Non-standardised	1	**Size of sample of speech and language delayed children**	
Criterion-referenced	1	50+	3
		25–49	2
		< 25	1
		Geographical location	
		More than one town/city	1
		No	0
		Areas of language investigated	
		Two or more	1
		One	0
		Duration of follow-up interval	
		> 3 years	3
		1–3 years	2
		> 6 months < 1 year	1
		Non-intervention	
		No speech and language intervention	4
Score		No specific speech and language intervention	1
Total		**Total**	
See summary table		**See summary table**	

Study reliability *(Quality of reporting):*

Study validity *(Internal and external):*

C) EVALUATION OF EFFECTIVENESS OF INTERVENTION: RCT AND QUASI-EXPERIMENTAL DESIGNS

REVIEWER: (Initials)

DATE OF CODING:

A. STUDY DETAILS

TITLE:

AUTHOR(S):

SOURCE_NAME AND REF:

SOURCE_DATE:

B. SUBJECT CHARACTERISTICS

Age range *(Months):*

Mean *(Or median, if reported instead)* **age** *(Months – round to nearest month):*

SD *(Months):*

Size of sample:

Rationale for sample size *(Free text; this indicates if statistical power of the study was considered):*

Sampling/recruitment procedures

(3) Probability (or representative) random sampling *(i.e. sampled from the population using simple random sampling, systematic sampling, stratified random sampling, or cluster sampling: note this procedure requires the researcher to have access to a list of the total population)*

(2) Non-probability (i.e. non-random) sampling *(e.g. quota sampling, convenience sampling)*

(1) Selected *(e.g. selected not sampled from group of diagnosed children or where professional nominates a proportion of children who attend a routine clinic without knowledge of case status)*

(0) Procedures are insufficiently recorded.

% Take-up rate *(Indicates the % of children sampled who actually participate in the study):*

Inclusion/exclusion criteria *(Free text: note any information provided by the author(s):*

No. of any excluded from sample *(Indicate % excluded due to any of the following):*

(1) Developmental delay (2) ESL (3) Other (specify)

Co-existing disabilities *(Indicate any co-existing disabilities and the % balance: disabilities/total):*

(1) Neurodevelopmental immaturities (e.g. hand–eye coordination):

(2) Specific learning difficulties:

(3) Medical complications (e.g. conductive hearing impairment: specify):

(4) Behaviour:

(5) Others (specify):

Description of sample

SES *(Indicate here % balance of SES [non-manual/total], etc., e.g. 23% of sample non-manual):*

Ethnicity *(Indicate here % balance ethnicity [ethnic groups/total], e.g. 17% of sample black):*

Gender *(Indicate here % balance gender [males/total], e.g. 56% of sample male):*

Geographical location *(Indicate here the country in which the study was carried out.*

Indicate below whether sample is from one town/area, etc.)

(1) One county/health authority: *(i.e. more than one town/city)*

(2) Sample from one town/city

(3) Other (specify)

Urban *(Indicate here % balance of urban, e.g. urban/total):*

Information about aetiology *(Free text):*

C. DESIGN

Intervention *(Note whether subjects are randomly allocated to groups [and the units of randomisation: individuals/classes/groups] and whether there was any matching)*

I. Randomised controlled trials (pre-test–post-test, with randomly allocated no treatment control group) (experimental)
Note units of randomisation (individuals/classes/groups)

II. Non-equivalent (pseudo-random/non-random no treatment control group (pre-test–post-test) (quasi-experimental) (including multiple time series with non-equivalent control)

Description of any matching between treatment and control groups *(Free text):*

Basis for any matching *(e.g. on pre-test score, IQ, etc. Free text):*

D. AREAS OF INVESTIGATION
Indicate using the table below (e.g. ***attention control/concentration/listening skills, prelinguistic skills*** *[such as turn-taking, babbling, eye contact, concept of object permanence, development of visual and auditory attention, etc.];* ***speech*** *[including articulation, phonology, motor planning/praxis, etc.];* ***expressive language*** *[including morphology, syntax, vocabulary, other aspects of semantics, etc.];* ***receptive language*** *[including syntax, vocabulary, understanding of concepts, etc.];* ***general language*** *[area unspecified];* ***pragmatics/social use of language, cognitive abilities*** *[e.g. IQ, memory] and* ***parent–child interaction****. Note that these areas are not mutually exclusive: indicate as many as are relevant).*

Area of language	**Investigated (Yes/No)**
Attention control/concentration/listening skills	
Prelinguistic skills	
Speech	
Expressive language	
Receptive language	
Expressive + receptive language	
Pragmatics	
General language (unspecified)	
Cognitive abilities (including IQ, memory, etc.)	
Parent–child interaction	
Other (Specify)	

E. STUDY CHARACTERISTICS *(Information here addresses issues relating to possible bias which may affect construct validity)*

Who carried out assessments? *(i.e. was it the author(s) or someone independent? Was the assessor 'blind' to the status of the children seen?):*

Who carried out the interventions? *(i.e. the author(s) or someone independent?):*

Setting (Clinic/hospital/laboratory/school/nursery/home):

Other (Specify):

Intervenor:
(a) Researcher/clinician (b) Parent(s) (c) Teachers (d) Peers

Description of treatment

Frequency of intervention session	
Length of sessions	
Duration of programme	
Treatment fidelity *(Free text: note any comments)*	

F. OUTCOME FINDINGS

Outcome measures *(Indicate the range of output measures used: more than one category may be used, if multiple measures were used):*

IQ Expressive language Receptive language Functional communication

Cognitive skills Social skills Academic skills

Area covered *(If WHO definitions are used):*

Impairment Disability Handicap Distress/well-being

Instrumentation *(Type of measures used):*

Norm-referenced Criterion-referenced Qualitative ratings

Measures used
(Complete tables below. Include name of any test[s]/variables; if multiple outcome measures, indicate all)

Treatment group(s)
Treatment 1: No. in group
Treatment 2: No. in group
Treatment 3: No. in group

Area and name of test(s) used or other outcome measures *(Specify)*	**Pre-test**		**Post-test**	
	Mean/median *(Indicate which)*	**SD**	**Mean/median** (Indicate which)	**SD**

Control group(s) *(Must be non-intervention)*
Control 1: No. in group
Control 2: No. in group

Area and name of test(s) used or other outcome measures *(Specify)*	**Pre-test**		**Post-test**	
	Mean/median *(Indicate which)*	**SD**	**Mean/median** (Indicate which)	**SD**

Details of any statistical tests used *(Insert appropriate values from study or use a separate sheet, if preferred):*

t F Chi-square *p*
% Frequency counts Effect size (from DSTAT) Corrected effect size

G. STUDY QUALITY RATINGS

Study reliability *(Factors relating to replicability of the study: score information provided as indicated below)*	**Score**	**Study validity** *(Factors relating to internal and external validity: score information provided as indicated below)*	**Score**
Subject		**Sampling Recruitment**	
Details of age:		Random	3
Mean	1	Non-random	2
Range	1	Selected	1
SD	1	Not given	0
Details of SES	1		
Details of gender	1	**Individual assignment to groups**	
Details of ethnicity	1	Random	3
Details of basis for subject classification		Non-random	2
(e.g. as language-delayed, etc.)	1	Selected only	1
		Not given	0
Materials			
Description, list or names of materials used	1	**Attrition**	
		< 30%:	2
Procedures		> 30%:	
Details of examiner/therapist	1	Replaced or controlled or intention-to-treat analysis	1
Directions for administration	1	Nothing done	0
Scoring	1		
		Inclusion/exclusion criteria	
Treatment		Given	1
Number of sessions per week	1	Not given	0
Length of sessions	1		
Duration of programme	1	**Definition of subjects' speech and language problem**	
Target area(s)/skill(s)	1	Given	1
Who measures outcomes	1	Not given	0
		Use of multiple outcome measures per area	
Instrumentation		Yes	1
Standardised	1	No	0
Criterion-referenced	1		
Qualitative ratings	1	**Blinding of assessments**	
		Yes	1
		No	0
		Comparability of pre-test scores (i.e. $p > 0.1$)	3
		Pre-test control ***(if pre-test scores are not comparable)***	
		Ss matched on pre-test scores	2
		Statistical control (e.g. ANCOVA)	1
		Not given	0
Total		**Total**	

Study replicability *(Quality of reporting):* **Study validity** *(Internal and external):*

Cost-effectiveness *(Insert any information regarding cost-effectiveness. Free text):*

Components of treatment regimen considered by author to have most impact on intervention result *(Free text):*

D) EVALUATION OF EFFECTIVENESS OF INTERVENTION: SINGLE-SUBJECT EXPERIMENTAL DESIGNS

REVIEWER: DATE OF CODING:
(Initials)

A. STUDY DETAILS

TITLE:

AUTHOR(S):

SOURCE_NAME AND REF: SOURCE_DATE:

B. SUBJECT CHARACTERISTICS

Age range *(Months):*

Mean *(Or median, if reported instead)* **age** *(Months – round to nearest month):*

SD *(Months, if n > 3):*

Size of sample:

Sampling/recruitment procedures
(3) Probability (or representative) random sampling *(i.e. sampled from the population using simple random sampling, systematic sampling, stratified random sampling, or cluster sampling. Note this procedure requires the researcher to have access to a list of the total population)*

(2) Non-probability (i.e. non-random) sampling *(e.g. quota sampling, convenience sampling)*

(1) Selected *(e.g. selected not sampled from group of diagnosed children or where professional nominates a proportion of children who attend a routine clinic without knowledge of case status)*

(0) Procedures are insufficiently recorded.

Take-up rate *(Indicates the number of children sampled who actually participate in the study):*

Inclusion/exclusion criteria *(Free text: note any information provided by the author(s):*

No. excluded from sample *(Indicate number excluded due to any of the following):*

(1) Developmental delay (2) ESL (3) Other (specify)

Co-existing disabilities *(Indicate any co-existing disabilities and the % balance: disabilities/total):*

(1) Neurodevelopmental immaturities (e.g. hand–eye coordination):

(2) Specific learning difficulties:

(3) Medical complications (e.g. conductive hearing impairment: specify):

(4) Behaviour:

(5) Others (specify)

Description of sample

SES *(Note any information for individual subjects):*

Ethnicity *(Note any information for individual subjects):*

Gender *(Note any information for individual subjects):*

Geographical location *(Indicate here the country in which the study was carried out.*

Indicate below whether sample is from one town/area, etc.):

(1) One county/health authority (i.e. more than one town/city)

(2) Sample from one town/city

(3) Other (specify)

Urban *(Indicate here % balance of urban (e.g. urban/total):*

Information about aetiology *(Free text):*

C. DESIGN

Intervention

I. Withdrawal and reversal designs
- (a) ABAB
- (b) ABA

II. Multiple baseline designs
- (a) across behaviours
- (b) across settings
- (c) across subjects
- (d) multiple probes variant

III. Alternating treatment designs

D. AREAS OF INVESTIGATION

Indicate using the table below (e.g. ***attention control/concentration/listening skills, prelinguistic skills*** *[such as turn-taking, babbling, eye contact, concept of object permanence, development of visual and auditory attention etc.];* ***speech*** *[including articulation, phonology, motor planning/praxis etc.];* ***expressive language*** *[including morphology, syntax, vocabulary, other aspects of semantics etc.];* ***receptive language*** *[including syntax, vocabulary, understanding of concepts etc.];* ***general language*** *[area unspecified];* ***pragmatics/social use of language, cognitive abilities*** *[e.g. IQ, memory] and* ***parent–child interaction****. Note that these areas are not mutually exclusive: indicate as many as are relevant).*

Area of language	Investigated (Yes/No)
Attention control/concentration/listening skills	
Prelinguistic skills	
Speech	
Expressive language	
Receptive language	
Expressive + receptive language	
Pragmatics	
General language (unspecified)	
Cognitive abilities (including IQ, memory, etc.)	
Parent–child interaction	
Other (Specify)	

E. STUDY CHARACTERISTICS *(Information here addresses issues relating to possible bias which may affect construct validity)*

Who carried out assessments? *(i.e. was it the researcher or therapist, or someone independent? Was the assessor 'blind' to the status of the children seen?):*

Setting *(Clinic/hospital/laboratory/school/nursery/home):*

Other *(specify)*

Intervenor

(a) Researcher/clinician (b) Parent(s) (c) Teachers (d) Peers

Intervention model used *(If more than one treatment, note also the most effective):*

(1) Didactic approaches (e.g. direct elicitation, modelling + prompting, imitation, mand-model approach, etc.)

(2) Naturalistic model only (e.g. using only naturalistic interaction in naturally occurring contexts)

(3) Hybrid (i.e. elicitation + natural conversational approaches) (e.g. milieu and extended milieu)

(4) Other (e.g. comprehension monitoring, etc.)

Details of treatment/intervention

Length of sessions:

Frequency of sessions (e.g. how many per week):

Duration of treatment/intervention:

Reliability of coding *(Note details of inter-rater/intra-rater reliabilities):*

Details of treatment/intervention (cont)

Baseline *(Details, free text):*

Intervention *(Details, free text):*

1.

2.

3.

4.

Maintenance *(Details, free text):*

Generalisation *(Details, free text):*

Follow-up *(Details, free text):*

Range of baseline sessions across subjects *(i.e. no. of baseline sessions for each subject):*	Range of treatment sessions across subjects *(i.e. no. of treatment sessions for each subject)* Range of maintenance sessions across subjects *(i.e. no. of maintenance sessions for each subject)*	Details of any untreated control processes *(additional experimental control) (specify):*
Criterion for success *(i.e. criterion for concluding phase of intervention or treatment: e.g. 90% success, etc.)*		Details of any reinforcement procedures *(e.g. rewards:social rewards, tangible rewards (non-edible), and edible rewards; time delays)*
	Generalisation training: (0) none *(e.g. no training for generalisation provided: 'train and hope' approach)* (1) general/specific *(e.g. training for generalisation provided, either general (i.e. unsystematic) or specific (i.e. systematic) procedure)*	Generalisation effects: *(does treatment outcome generalise to another setting, e.g. natural conversation, or to another behaviour, or to other stimuli or to a combination of stimuli and settings)* Setting Behaviour Stimuli Stimuli and setting

F. OUTCOME FINDINGS *(Indicate the range of output measures used: more than one category may be used, if multiple measures were used):*

IQ Expressive language Receptive language Functional communication

Cognitive skills Social skills Academic skills

Area covered *(If WHO definitions are used):*

Impairment Disability Handicap Distress/well-being

Instrumentation *(Type of measures used):*

Norm-referenced Criterion-referenced Qualitative ratings

Dependent variables:
(Complete table below. Include name of any test[s]/variables; if multiple outcome measures, indicate all. If there is more than one outcome measure per subject for a given treatment, generalisation or follow-up outcome effect, calculate and report the median for each subject.)

Dependent variables *(specify)*	**Treatment outcome effect** *PND compared with baseline PND for each subject*	**Generalisation outcome effect** *PND compared with baseline PND for each subject*	**Follow-up outcome effect** *PND compared with baseline PND for each subject*

Summary *(Free text; note any individual variation)*

Details of any statistical tests used *(Insert appropriate values):*

G. STUDY QUALITY RATINGS

Study reliability *(Factors relating to replicability of the study: score information provided as indicated below)*		**Study validity** *(Factors relating to internal and external validity: score information provided as indicated below)*	
	Score		**Score**
Subject		**Subject selection**	
Details of age:		Random sampling	3
Range	1	Non-random sampling	2
Details of SES	1	Selected	1
Details of gender	1	Not given	0
Details of ethnicity	1		
Details of basis for subject classification (e.g. as language-delayed, etc.)	1	**Inclusion/exclusion criteria**	
		Given	1
Materials		Not given	0
Description, list or names of materials used	1	**Definition of subject's speech and language problem**	
		Given	1
Procedures		Not given	0
Details of examiner/therapist	1	**Use of untreated control processes**	
Directions for administration	1	Yes	2
Scoring details	1	No	0
		Median length of baseline for study	
		≥ 7 sessions	2
		< 7 sessions	0
		Stability of baseline	
		For all Ss: ± 10% of mean probe score for each subject	2
		For any S: outwith ± 10% of mean probe score for each subject	0
		Median length of treatment phase for study	
		≥ 18 sessions	2
		< 18 sessions	0
		Generalisation phase	
		Yes	2
		No	0
		Replication (i.e. n > 1)	
		Yes	3
		No	0
Total		**Total**	

Cost-effectiveness *(Insert any information regarding cost-effectiveness e.g. parents or teachers as intervenors, etc. Free text):*

Components of treatment regime considered by author to have most impact on intervention result *(Free text):*

E) EVALUATION OF EFFECTIVENESS OF SCREENING TESTS

(Notes for coders are in italics)

REVIEWER: DATE OF CODING:
(Initials)

A. STUDY DETAILS

TITLE:

AUTHOR(S):

SOURCE_NAME AND REF: SOURCE_DATE:

PART OF A SERIES OF PAPERS ON THIS TEST (Y/N/NK) *(This is a check for multiple or related reporting about the same cohort)*

B. SCREENING PROCEDURE

Name of test:

Areas tested:

-
-
-
-

Rec. age *(Months):*

Admin. time *(Minutes):*

Sources of information *(Use 1–5):*

Methods *(Use 1–4):*

Mono/multiphasic *(Insert 1 or 2):*

Any information on costs? *(Y or N; if Yes then note details below):*

C. CRITERIA FOR DELAY

Name of reference test or description of procedure	Areas tested	Cut-off criteria for delay	Rationale for cut-off

D. SAMPLE

Sampling/recruitment procedures *(Use 0–3):*

Sample descriptions

	Original sample	Screened sample	Concurrent sample	Predictive sample
Nature of sample Normative/clinical/mixed				
Number				
Exclusions				
SES *(Indicate here % balance of SES [non-manual/total], etc., e.g. 23% of sample non-manual)*				
Ethnicity *(Indicate here % balance ethnicity [ethnic groups/ total] e.g. 17% of sample black)*				
Gender *(Indicate here % gender balance [males/total) e.g. 56% of sample male)*				
Age range *(Months)*				
Mean age *(Or median, if reported instead; months – round to nearest month)*				
SD age *(Months)*				

Nature of predictive sample

(a) randomly selected
(b) self-selected/skewed
(c) no details available

Number in screen sample having full diagnostic testing:

and as % of screened sample:

How was sample chosen for full diagnostic testing? *(Use 1–3):*

Geographical location
(Indicate here the country in which the study was carried out.

Indicate below whether sample is from one town/area, etc.)

(1) One county/health authority: i.e. more than one town/city

(2) Sample from one town/city

(3) Other (specify)

Urban *(Indicate here % balance of urban (e.g. urban/total):*

E. RELIABILITY *(Record details of any reliability coefficients quoted in the study for the screening procedure and also grade using the 1–3 scale provided):*

(3) High ($r \geq 0.8$)

(2) Moderate ($r \geq 0.5, < 0.8$)

(1) Low ($r < 0.5$)

Inter-rater:
Rating: 1 2 3

Test–re-test:
Rating: 1 2 3

F. VALIDITY *(Record details of any validity measures quoted in the study: complete the 2 × 2 table and calculate as stated)*

Concurrent

	Reference test positive*	Reference test negative	**Total**
Screen positive*	a	b	
Screen negative	c	d	
Total	a + c	b + d	

* Positive refers to presence of disorder

Predictive

	Reference test positive*	Reference test negative	**Total**
Screen positive*	a	b	
Screen negative	c	d	
Total	a + c	b + d	

* Positive refers to presence of disorder

Summary

	Concurrent	**Rate** h/m/l*	**Predictive**	**Rate** h/m/l*
Sensitivity a / a + c				
Specificity d / b + d				
Positive predictive value a / a + b				
Relative risk [a / a + b] / [c / c + d]				
Likelihood ratio [a / a + c] / [b / b + d]		Not applicable		Not applicable
Odds ratio ad/bc		Not applicable		Not applicable

* (3) high ($r \geq 0.8$), (2) moderate ($r \geq 0.5, < 0.8$), (1) low ($r < 0.5$)

Other output measures
(These may be supplied in related papers: give reference to work and details of validity figures here)

Reference:
Construct validity:
Content validity:
Other:

G. QUALITY RATINGS

Study reliability *(Factors relating to replicability of the study: score information provided as indicated below)*		**Study validity** *(Factors relating to internal and external validity: score information provided as indicated below)*	
	Score		**Score**
Subject		**Publication**	
Details of age:		Sampling procedure:	
Report one of: mean, range or SD	1	0–3 as in form above	
Details of gender:	1	0	0
Report details of one of: ethnicity, SES	1	1	1
		2	2
Procedures		3	5
For the screening test:			
Details of the cut-off	2	Sample from general population	
Rationale for the cut-off	2	(mixed or clinical gets 0)	0 or 2
Details of administration (reference to a published test acceptable)	2		
		Sample size given diagnostic testing:	
		50/150/150+	1–3
For the reference test		**Reference testing**	
Details of the cut-off	2	Reference test valid	1
Rationale for cut-off	2	i.e. published test or clinical judgement objectified	
For the use of clinical judgement		Choice of children to give diagnostic test to:	
Description of process	1	Only screen fails	1
Method given for reaching decision	1	Selected screen fails and some screen passes (not specified or not stratified selection)	2
Reliability		Selected screen fails and passes-	
Details of test–retest; interrater reliability	1 or 2	specified as stratified or random	3
		All of sample screened	3
		Independent assessors or random order of tests	1
		Later tester blind to results of earlier test	1
Score out of 17:		Score out of 16:	
Grade low: **Grade medium:** **Grade high:**		**Score summary: ____of 17 _____of 16 _____of 33**	

Appendix 9

Intervention studies: quality

TABLE 37 *Study quality of RCT experimental designs*

Study (reliability and validity scores)*	Country	Areas of language treated†	Sampling/ recruitment procedures	Assignment of subjects to groups	Blinding of assessors to subject status	Comparability of pre-test scores ($p > 0.1$)
Almost and Rosenbaum, 1998 (R15; V13)	Canada	1	Selected	Random	Yes	Yes
Fey *et al,* 1993 (R15; V12)	USA	2	Non-random + selected	Random	Yes	Yes
Gibbard, 1994 (Study 1) (R16; V12)	UK	2	Selected	Random	Not given‡	Yes
Girolametto *et al,* 1995 (R14; V12)	Canada	2	Non-random	Random	Yes	Yes¶
Girolametto *et al,* 1996 (R16; V14)	Canada	2, 6	Non-random	Random	Yes	No
Lancaster, 1991 (R15; V11)	UK	1	Selected	Random	Not given	Yes
Methany and Panagos, 1978 (R11; V12)	USA	1, 2	Selected	Random	Not given	Yes
Reid *et al,* 1996 (R12; V12)	UK	1	Selected	Random	Not given	Insufficient data reported
Schwartz *et al,* 1985 (R13; V9)	USA	2	Selected	Random	Not given	No
Shelton *et al,* (Study 1) (R14; V10)	USA	1, 3, 4	Non-random	Random	Not given	Yes**

* *Study reliability (R) scores are out of a maximum of 19 and study validity (V) scores have a maximum of 15.*

† *1 = phonology/articulation, 2 = expressive language (syntax and/or vocabulary), 3 = receptive language (comprehension and/or vocabulary), 4 = auditory discrimination/listening skills/phonemic awareness, 5 = pragmatics, and 6 = parent–child interaction.*

‡ *Language sample measures were coded 'blind'.*

¶ *Pre-test scores on language measures were comparable but between-group differences were found for pre-test scores on a behaviour measure.*

** *There was a significant difference between the two groups in pre-test scores for the 'noise' condition of the Goldman-Fristoe-Woodcock Test of Auditory Discrimination (Goldman, Fristoe and Woodcock, 1970).*

TABLE 38 *Study quality of quasi-experimental designs*

Study (reliability and validity scores)*	Country	Areas of language treated†	Sampling/ recruitment procedures	Assignment of subjects to groups	Blinding of assessors to subject status	Comparability of pre-test scores
Conant *et al,* 1984 (R14; V9)	USA	2, 5	Selected	Non-random	Not given	No
Fey *et al,* 1994 (R17; V11)	USA	1,2	Selected	Non-random (for immediate vs. delayed groups)	Not given	Yes
Gibbard, 1994 (Study 2) (R15; V9)	UK	2	Selected	Non-random	Not given	No‡
McDade and McCartan, 1996 (R19; V12)	UK	2, 3, 6	Selected	Non-random	Not given¶	Yes
Shelton *et al,* 1978 (Study 2) (R10; V8)	USA	1	Selected	Non-random	Not given	Yes**
Stevenson *et al,* 1982 (R16; V13)	UK	2 ,3	Selected	Non-random	Yes	Yes
Ward, 1994 (Group 1) (R13; V10)	UK	2, 3	Selected	Non-random	Not given	Yes
Ward, 1994 (Group 2) (R13; V10)	UK	2, 3	Selected	Non-random	Not given	Yes
Warrick *et al,* 1993 (Study 2) (R13; V10)	USA	4	Selected	Non-random	Not given	Yes
Whitehurst *et al,* 1991 (R13; V6)	USA	2	Selected	Non-random	Not given	Insufficient data reported
Wilcox and Leonard, 1978 (R9; V10)	USA	2	Selected	Non-random	Not given	Yes
Zwitman and Sonderman, 1979 (R10; V9)	USA	2	Selected	Non-random	Not given	Yes

* *Study reliability (R) scores are out of a maximum of 19 and validity (V) scores out of a maximum of 15 for each study.*

† *1 = phonology/articulation, 2 = expressive language (syntax and/or vocabulary), 3 = receptive language (comprehension and/or vocabulary and/or auditory association), 4 = auditory discrimination/listening skills/phoneme awareness, 5 = pragmatics, and 6 = parent–child interaction.*

‡ *There were significant between-group differences on the Reynell Expressive Language pre-test scores.*

¶ *An independent therapist carried out some of the post-assessment along with one of the authors. However, no information is provided as to whether the assessors were aware of the treatment status of the children they saw.*

** *However, there was marked variability in pre-test scores and only four to five subjects in each of the groups.*

TABLE 39 *Study quality of single-subject experimental designs: higher quality studies (study validity scores ≥ 10)*

Study (n = no. of subjects)	Areas of language treated and design*†	Untreated control processes	No. baseline sessions (range)	Stability of baselines (± 10% of mean baseline score)	No. treatment sessions (range)	Generalisation	n > 1
Bedrosian and Willis, 1987 (n = 1)	5 (c)	Yes	3–36	1/3 stable	11–33	No	No
Camarata, 1993 (n = 2)	1 (a, c)	No	3–15	5/5 stable	8–24	Yes‡	Yes
Connell, 1986a (n = 2)	2 (a)	No	5–11	2/2 stable	11–17	Yes¶	Yes
Connell, 1986b (n = 4)	2 (a)	No	10–20	2/4 stable	20–70	Yes¶	Yes
Ellis Weismer *et al,* 1993 (n = 3)	2 (e)	No	4	4/4 stable	24	No	Yes
Gierut, 1990 (n = 3)	1 (a, e)	No	4–8	5/6 stable	22	Yes**	Yes
Hegde and Gierut, 1979 (n = 1)	2 (c)	No	10–40	4/4 stable	20–45	Yes**	No
Hegde *et al,* 1979 (Study 2) (n = 1)	2 (c)	No	15–40	4/4 stable	35–65	Yes**	No
Hemmeter and Kaiser, 1994 (n = 1)	2, 3, 6 (d)	No	23	2/3 stable	21	Yes††	Yes‡‡
McGregor, 1994 (n = 2)	2 (a, c)	No	4–6	1/16 stable	18	Yes**	Yes
Olswang and Bain, 1985 (n = 3)	1 (a, f)	No	6–24	12/16 stable	4–24	Yes¶	Yes
Olswang *et al,* 1986 (n = 2)	2 (e)	No	9	2/4 stable	24–27	Yes¶¶	Yes
Olswang and Coggins, 1984 (n = 2)	2 (c)	No	3–35	3/12 stable	8–35	No	Yes
Powell *et al,* 1991 (n = 6)	1 (c)	Yes	3–12	18/28 stable	4–31	Yes**	Yes
Warren *et al,* 1984 (n = 3)	2, 6 (a)	No	9–18	0/3 stable	15–36	Yes¶¶	Yes
Williams, 1991 (n = 9)	1 (c)	No	9	13/18 stable	9–42	Yes**	Yes
Young, 1987 (n = 2)	1 (c)	Yes	3–23	2/6 stable	6–13	Yes**	Yes

* *1 = phonology/articulation, 2 = expressive language (syntax, semantics and/or vocabulary), 3 = receptive language (comprehension and/or vocabulary and/or auditory association), 4 = auditory discrimination/listening skills/phoneme awareness, 5 = pragmatics, and 6 = parent–child or teacher–child interaction.*

† *a = multiple baselines across subjects design, b = multiple baselines across groups, c = multiple baselines across behaviours, d = multiple probes, e = alternating treatments, and f = withdrawal.*

‡ *Generalisation to home settings.*

¶ *Generalisation to untrained probes and to spontaneous use in conversation.*

** *Generalisation to untrained probes.*

†† *Generalisation to a clinic setting with an unfamiliar adult and to the home setting.*

‡‡ *Three other subjects were excluded from analysis here on account of secondary delay.*

¶¶ *Generalisation to spontaneous production of untrained probes in free play setting.*

TABLE 40 *Study quality of single-subject experimental designs: lower quality studies (study validity scores < 10)*

Study (n = no. of subjects)	Areas of language treated and design*†	Untreated control processes	No. baseline sessions (range)	Stability of baselines (± 10% of mean baseline score)	No. treatment sessions (range)	Generalisation	n > 1
Culatta and Horn, 1987 (n = 3)	2 (c)	No	6	2/9 stable	8–13	Yes‡	Yes
Dollaghan and Kaston, 1986 (n = 2)	3 (a)	No	3–6	4/8 stable	11	Yes¶	Yes
Ellis Weismer and Murray-Branch, 1989 (n = 4)	2 (e)	No	4	1/4 stable	10–14	Yes¶	Yes
Gierut, 1992 (Study 1) (n = 4)	1 (a, e)	No	2–10	1/6 stable	8–20	Yes¶	Yes
Hargrove *et al,* 1989 (n = 1)	1 (c)	No	3–16	1/3 stable	5–7	Yes¶	No
Kaiser *et al,* 1995a (n = 2)	2, 6 (a)	No	8–10	0/2 stable	10–11	Yes**	Yes
Kaiser *et al,* 1995b (n = 2)	2, 6 (a)	No	4	1/4 stable	7–18	No	Yes
Olswang *et al,* 1983 (n = 3)	2 (e)	Yes	3-4	4/8 stable	9–20	No	Yes
Powell and Elbert, 1984 (n = 5)	1 (a)	No	3–7	0/5 stable	Not given††	Yes¶	Yes

* *1 = phonology/articulation, 2 = expressive language (syntax and/or vocabulary), 3 = receptive language (comprehension and/or vocabulary and/or auditory association), 4 = auditory discrimination/listening skills/phoneme awareness, 5 = pragmatics, and 6 = parent–child interaction.*

† *a = multiple baselines across subjects design, b = multiple baselines across groups, c = multiple baselines across behaviours, d = multiple probes, e = alternating treatments, and f = withdrawal.*

‡ *Generalisation to spontaneous production of targets.*

¶ *Generalisation to untrained probes.*

** *Generalisation to conversation and home setting.*

†† *Data are provided for probes presented at the end of Phase 3.*

Appendix 10

Intervention studies: outcomes and effect sizes

Descriptions of the different intervention approaches as classified here (see *Tables 41–52*) can be found in appendix 2. Note that intervention models are classified here in terms of their presentation to the child and unless otherwise stated the frequency of treatment refers to the clinician's contribution.

Details of the meta-analysis of effect sizes for RCT/quasi-experimental designs

Articulation/phonology

Four RCT and two quasi-experimental designs yielded 18 standardised effect sizes for outcomes following intervention for problems in articulation and/or phonology. *Table 45* provides a summary of the results. The table shows the effect sizes (*d*) for each study for direct and indirect treatment for norm-referenced tests and criterion-referenced measures across two different treatments in the case of two studies. Asterisks indicate that the observed effect size represents a statistically significant treatment effect (i.e. that the post-test scores from the treated group are significantly higher than those from the untreated control group).

The results reveal significant outcomes in terms of post-test comparisons from the Almost and Rosenbaum study. Analyses of variance were carried out on *d* values weighted by the inverse of their variance (Hedges and Olkin, 1985) using the *Q*-statistics for both norm-referenced and criterion-referenced measures for direct versus indirect service delivery. The results are summarised in *Table 46* for the 'best' treatment where there was more than one effect size for a study in one of the cells above.

The results from these four studies reveal that the outcomes from direct treatment were significantly higher than the effects from indirect, parent-administered treatment ($Qb(1) = 3.92$, $p < 0.05$).[1] However, caution is required in interpreting these results; in addition to the small number of studies in the comparison, the direct/indirect treatment variable is confounded with study design quality. That is, both of the effects from direct treatment come from RCTs while only one of the indirect treatment effects is from an RCT, and hence the difference between direct and indirect treatment above could be due to a difference in outcomes between RCTs and quasi-experimental designs.

However, further evidence is provided by the criterion-referenced measures (*Table 47*). Effect sizes are available from seven studies (four RCTs and three quasi-experimental designs) and the direct/indirect treatment variable is not confounded with study design. ANOVA (Hedges and Olkin, 1985) revealed no significant differences within the categories of direct and indirect treatment but a significant difference between the categories in favour of direct clinical treatment ($Qb(1) = 6.29$, $p < 0.02$).[2]

Note, however, the small sample sizes and the fact that the studies by Shelton and co-workers used parent-administered listening training as the sole treatment approach for articulation problems.

Expressive language

Six RCT and eight quasi-experimental designs yielded outcomes for intervention in expressive language. *Table 48* provides a summary of the results. Direct treatment yielded only one significant outcome, but eight for indirect treatment.

ANOVAs were carried out as before on *d* values weighted by the inverse of their variance (Hedges and Olkin, 1985) using the *Q*-statistics for norm-referenced and criterion-referenced measures separately for each study design and also for direct versus indirect service delivery. The results are summarised in *Table 49*.

[1] The equivalent *Qb* for the least successful treatment outcomes from the Methany and Panagos study and the study by Shelton and co-workers is 8.48, $p < 0.01$.

[2] A similar difference was observed in the case of the 'worst' treatment outcomes ($Qb[1] = 8.88$, $p < 0.01$).

The results for effect sizes based upon norm-referenced test scores firstly reveal that there are no significant differences between the effect sizes for RCTs and those for quasi-experimental studies, though the *Qw* statistic for RCTs only just meets the criterion for homogeneity. The data indicate a strong statistically significant treatment effect, indicating that the children on average made progress of around 1 SD on norm-referenced tests. For a child scoring at the 5th percentile on a standardised test with a SD of 15, progress of the order of 1 SD as a result of intervention would represent a shift to the 25th percentile (i.e. to within the normal range).[3]

A further analysis of direct versus indirect treatment revealed a marginally significant difference in favour of indirect treatment (*Table 50*). The *Qw* statistic reveals marked heterogeneity in the indirect treatment studies. Analysis revealed that the Gibbard study (1994, Study 1), with an effect size of +2.38, was an outlier. Removal of this study from the analysis improved the homogeneity of the effect sizes ($Qw(12) = 13.48$, $p = 0.34$) but the difference between the two groups remained marginal ($Qb(1) = 3.14$, $p = 0.08$).

The results from the analysis of the effect sizes based upon the criterion-referenced scores revealed a similar picture, with no difference between the two groups ($Qb(1) = 0.021$, $p = 0.88$).

While there is no significant difference between the two groups in terms of effect size for interventions in expressive language, the overall results indicate that indirect parent-administered treatment is at least as effective as direct clinical treatment in this area.

Receptive language

Two RCT and four quasi-experimental designs yielded effect sizes for receptive language. *Table 51* provides a summary of the results. The table shows the effect sizes for each study for direct and indirect treatment for norm-referenced tests and criterion-referenced measures, across two different treatments in the case of two studies. The asterisk indicates that the observed effect size represents a statistically significant treatment effect. The results reveal no significant treatment outcome for direct treatment but three for indirect treatment. ANOVAs were carried out as before on *d* values weighted by the inverse of their variance (Hedges and Olkin, 1985) using the *Q*-statistics for norm-referenced and criterion-referenced measures separately for study design and also for direct versus indirect service delivery. The results are summarised in *Table 52*.

The results indicate a significant difference between indirect and direct treatment, with indirect treatment resulting in highly significant effects of almost 1.5 SDs. However, caution is again required on account of the small number of studies and because of the marked heterogeneity in the sample of indirect treatment studies.

Auditory discrimination/listening skills/phoneme awareness

Three quasi-experimental studies were carried out in the broad area of auditory discrimination/listening skills. The two studies of parent-administered training in listening skills did not produce successful outcomes (Shelton *et al,* 1978, Studies 1 and 2), but the subjects who participated in the study by Warrick and co-workers (1993) phoneme training programme made significant gains ($d = +0.81$, $p < 0.05$).

[3] Only four studies actually carried out their primary analyses on standard scores but the average effects size from these studies weighted by the inverse of the variance of each study is +1.09 (95% CI, +0.77/+1.41), which is of the same order as the overall average.

TABLE 41 *Summary of treatment outcomes from studies using RCT designs*

Study/no. in groups (experimental and control)*	Area of intervention†	Child characteristics (mean age of sample/gender balance of sample/social class balance (% non-manual)	Programme characteristics (frequency and duration/ intervention approach)	Treatment outcomes
Almost and Rosenbaum, 1998 (13E, 13C)	1	42 months 81% male	Two sessions a week for 4 months (30 min/session). Didactic approach	Treated children made greater gains than controls ($p < 0.01$)
Fey *et al,* 1993 (11E1, 10E2, 9C)	2	56 months 70% male	Three sessions a week for 5 months (60 min/session) or 12×2 hours of group training for parents. Hybrid approach	Treated children made greater gains than controls ($p < 0.01$). No difference between direct treatment and parent-administered treatment
Gibbard, 1994 (Study 1) (18E, 18C)	2	35 months 66% male 66% non-manual	One session a fortnight for parents for 6 months (> 60 min/session). Hybrid approach	Parent-administered treatment led to greater gains than controls ($p < 0.01$)
Girolametto *et al,* 1996 (12E, 13C)	2, 6	29 months	One session a week for parents for 3 months (> 60 min/session). Hybrid approach	Parent-administered treatment led to greater gains than controls ($p < 0.05$)
Girolametto *et al,* 1995 (8E, 8C)	2	29 months	One session a week for 2.5 months (> 60 min/session). Hybrid approach	Parent-administered treatment led to greater gains than controls ($p < 0.05$)
Lancaster, 1991 (5E1, 5E2, 5C)	1	43 months 80% male	One or two sessions a week for 6 months (30–40 min/session) or 4 hours of group training for parents. Didactic approach	Treated children made greater gains than controls ($p < 0.05$). No difference between direct treatment and parent-administered treatment
Methany and Panagos, 1978 (8E1, 8E2, 8C)	1, 2	N/A	N/A for 5 months. Didactic approach	Treated children made greater gains than controls ($p < 0.05$). Transfer of training in syntax to phonology and vice versa
Reid *et al,* 1996 (8E1, 8E2, 8C)	1	N/A	One session a week for 1.5 months (30 min/session) versus one session a week for 2.5 months. Hybrid approach	Only children treated for 10 weeks made progress ($p < 0.05$)
Schwartz *et al,* 1985 (8E, 2C)	2	35 months 100% male	Three sessions a week for 1 month. Hybrid approach	Treated children increased scores ($p < 0.05$)
Shelton *et al,* (Study 1) (15E1, 15E2, 15C)	1, 3, 4	48 months treated subjects and 39 months controls	Five sessions a week from mother for 3 months (5–15 min/session). Didactic approach	Auditory training administered by parents did not improve sound production ($p > 0.1$)

* *E = experimental, C = control*

† *1 = phonology/articulation, 2 = expressive language (syntax and/or vocabulary), 3 = receptive language (comprehension and/or vocabulary and/or auditory association), 4 = auditory discrimination/listening skills/phoneme awareness, 5 = pragmatics, and 6 = parent–child interaction*

N/A = not available

continued

TABLE 42 *Summary of treatment outcomes from studies using quasi-experimental designs*

Study/no. in groups (experimental and control)*	Area of intervention†	Child characteristics (mean age of sample/gender balance of sample/social class balance (% non-manual)	Programme characteristics (frequency and duration/ intervention approach)	Treatment outcomes
Conant *et al,* 1984 (26E, 22C)	2, 5	53 months	Two or three sessions a week for 4 months (30–45 min/session). Hybrid approach	Treated children with primary delay made greater gains than controls ($p < 0.05$) in syntax and pragmatic use of language
Fey *et al,* 1994 (10E1, 8E2, 8C)	1, 2	56 months 70% male	Three sessions a week for 5 months (60 min/session) or 12 x 2 hours of group training for parents. Hybrid approach	Treated children made greater gains than controls in syntax ($p < 0.001$) but training in syntax did not generalise to phonology. No difference between direct treatment and parent-administered treatment
Gibbard, 1994 (Study 2) (8E1, 9E2, 8C)	2	32 months 76% male 35% non-manual	One session a week for 6 months (30 min/session) or half sessions a week for 6 months (> 60 min/session). Hybrid approach	Parent-administered treatment resulted in greater gains than controls on all measures ($p < 0.01$), while clinician-treated children made significant gains on only two measures. Parent-administered treatment also led to longer utterances than direct clinician treatment ($p < 0.01$)
McDade and McCartan, 1996 (9E, 9C)	2, 3, 6	27 months	One session a week for 3 months (> 60 min/session). Naturalistic approach	Parent-administered treatment led to greater gains in expressive language than controls ($p < 0.01$)
Shelton *et al,* 1978 (Study 2) (4E1, 5E2, 5C)	1	51 months treated subjects and 42 months controls	Ten sessions (15 min/session) for an unspecified duration. Hybrid approach	Auditory training administered by parents did not improve sound production ($p > 0.1$)
Stevenson and Bax, 1982 (12E, 10C)	2, 3	36 months 55% male 0% non-manual	One session a week for 6 months. Didactic approach	Both treated and control children made gains in expressive language ($p < 0.05$); no difference between the two groups
Ward, 1994 (Study 1) (43E, 41C)	2, 3	10 months	One session a month (35 min/session) for 4 months. Hybrid approach	Treated children made greater gains than controls in both expressive and receptive language (no details of any statistical tests)
Ward, 1994 (Study 2) (9E, 4C)	1	10 months	One session a month (35 min/session) for 4 months. Hybrid approach	Treated children made greater gains than controls in both expressive and receptive language (no details of any statistical tests)

* *E = experimental, C = control*

† *1 = phonology/articulation, 2 = expressive language (syntax and/or vocabulary), 3 = receptive language (comprehension and/or vocabulary and/or auditory association), 4 = auditory discrimination/listening skills/phoneme awareness, 5 = pragmatics, and 6 = parent–child interaction*

N/A = not available

continued

TABLE 42 contd *Summary of treatment outcomes from studies using quasi-experimental designs*

Study/no. in groups (experimental and control)*	Area of intervention†	Child characteristics (mean age of sample/gender balance of sample/social class balance (% non-manual)	Programme characteristics (frequency and duration/ intervention approach)	Treatment outcomes
Warrick *et al,* 1993 (14E, 14C)	4	N/A	Two sessions a week for 2 months (20 min/session). Didactic approach	Treated children made greater gains than controls ($p < 0.05$). The advantage was still evident a year later ($p < 0.03$)
Whitehurst *et al,* 1991 (25E, 37C)	2	28 months	One session a fortnight for an unspecified period (30 min/session). Hybrid approach	Treated children achieved higher scores than controls ($p < 0.01$) and a higher proportion normalised their scores ($p < 0.02$)
Wilcox and Leonard, 1978 (12E, 12C)	2	68 months treated subjects and 58 months controls	N/A. Didactic approach	Treated children achieved higher scores than controls ($p < 0.001$)
Zwitman and Sonderman, 1979 (11E, 11C)	2	48 months	One session a week for 5 months (45 min/session). Didactic approach	Treated children made greater gains than controls ($p < 0.05$)

* *E = experimental, C = control*

† *1 = phonology/articulation, 2 = expressive language (syntax and/or vocabulary), 3 = receptive language (comprehension and/or vocabulary and/or auditory association), 4 = auditory discrimination/listening skills/phoneme awareness, 5 = pragmatics, and 6 = parent–child interaction*

N/A = not available

TABLE 43 *Summary of treatment outcomes for single-subject designs: higher quality studies (study validity scores ≥ 10)*

Study	Area of intervention*	Child characteristics†	Programme characteristics (frequency and duration/ intervention approach)	Treatment outcomes
Bedrosian and Willis, 1987	5	n = 1M (60 months)	Two sessions a week for 6 months (30 min/session). Hybrid approach	Criterion for success in future-related and memory-related initiations was achieved
Camarata, 1993	1	n = 2 (1M, 46 months and 1F, 51 months)	Two sessions a week for 2–4 months (45 min/session). Naturalistic approach	Children achieved criterion for production for target sounds. Gains generalised and were maintained 9 months later
Connell, 1986a	2	n = 2 (mean 36 months)	Three sessions a week for 1–2 months (40 min/session). Didactic approach	Children all reached the criterion for success for production of semantic roles and generalised use to spontaneous conversation
Connell, 1986b	2	n = 4 (2M, 2F, mean 45 months)	Three to four sessions a week for 2–6 months (30 min/session). Didactic approach	Children all reached the criterion for success for the subject properties taught and generalised use to spontaneous conversation
Ellis Weismer *et al,* 1993	2	n = 3 (2M, 1F, mean 27 months)	Two sessions a week for 3 months (60 min/session). Naturalistic vs. hybrid approach	One of the children made most gains in response to modelling, one to modelling plus evoked production and neither approach was effective in the case of the third
Gierut, 1990	1	n = 3M (mean 54 months)	Three sessions a week (60 min/session). Didactic approach	Treatment of maximal oppositions was most effective
Hegde and Gierut, 1979	2	n = 1M (57 months)	Four sessions a week for 3 months (30 min/session). Didactic approach	Children achieved the criterion for success for all four behaviours
Hegde *et al,* 1979 (Study 2)	2	n = 1M (48 months)	Four to five sessions a week for 3 months (45 min/session). Didactic approach	Child achieved the accuracy criterion for success for all features
Hemmeter and Kaiser, 1994	2, 3	n = 1F (25 months)	Two sessions a week for 5 months (45 min/session). Hybrid approach	Parent learned to use enhanced milieu techniques and child made gains which generalised to the home
McGregor, 1994	2	n = 2 (1M, 56 months, 1F, 60 months)	Two sessions a week for 3 months. Didactic approach	Phonologically-based treatment improved word finding

* *1 = phonology/articulation, 2 = expressive language (syntax, semantics and/or vocabulary), 3 = receptive language (comprehension and/or vocabulary and/or auditory association), 4 = auditory discrimination/listening skills/phoneme awareness, 5 = pragmatics, and 6 = parent–child interaction*

† *M = male, F = female*

TABLE 44 *Summary of treatment outcomes for single-subject designs: lower quality studies (study validity scores < 10)*

Study	Area of intervention*	Child characteristics†	Programme characteristics (frequency and duration/ intervention approach)	Treatment outcomes
Culatta and Horn, 1987	2	n = 3 (1M, 2F, mean 67 months)	Two sessions a week for 2–3 months (45 min/session). Hybrid approach	Training generalised to spontaneous productions and highlights the effectiveness of using meaningful communicative contexts
Dollaghan and Kaston, 1986	3	n = 2 (70 and 75 months)	Three sessions a week for 1 month (20 min/session). Comprehension monitoring approach	Treatment led to both children querying problem messages and generalised to different and more complex messages
Ellis Weismer and Murray-Branch, 1989	2	n = 4 (3M, 1F, mean 70 months)	One to two sessions a week for 2 months (20–25 min/session). Naturalistic vs. hybrid approach	Modelling plus evoked production tended to be associated with less variable learning curves
Gierut, 1992 (Study 1)	1	n = 4 (3M, 1F, mean 46 months)	Three sessions a week 1–1.5 months (60 min/session). Didactic approach	Teaching minimal pairs with two new phonemes unknown to the child is as effective or more effective than teaching one new phoneme
Hargrove *et al,* 1989	1	n = 1M (80 months)	Two sessions a day for 9 days. Didactic approach	Training in prosody skills improved performance but the length of the treatment phase might have been extended
Kaiser *et al,* 1995a	2	n = 2 (1M, 37 months, 1F, 43 months)	Two sessions a week for 1–1.5 months (30–60 min/session). Hybrid approach	Parents learned to use milieu teaching approaches and the children increased their use of targets. Only one of the children showed generalised use of targets to the home setting
Kaiser *et al,* 1995b	2	n = 2 (42 and 46 months)	Eight sessions (60–90 min/session). Hybrid approach	Both parents learned to use milieu teaching techniques. Variation in the relative effectiveness of individual parent training and individual parent training noted
Olswang *et al,* 1983	2	n = 3 (2M, 1F, range 23–40 months)	Three sessions a week for 2 months (30 min/session). Didactic approach	Picture identification was the most effective treatment for lexical learning for two of the children
Powell and Elbert, 1984	2	n = 5 (4M, 1F, range 52–67 months, mean 60 months)	30-minute sessions over a 9-month period, with treatment lasting for 2–4 months. Didactic approach	Both treatments (stop plus liquid and fricative plus liquid) generalised to treated and untreated cluster categories

* *1 = phonology/articulation, 2 = expressive language (syntax and/or vocabulary), 3 = receptive language (comprehension and/or vocabulary and/or auditory association), 4 = auditory discrimination/listening skills/phoneme awareness, 5 = pragmatics, and 6 = parent–child interaction*

† *M = male, F = female*

TABLE 45 *Summary of effect sizes for articulation/phonology outcome measures*

Study	Effect size and 95% CI for direct treatment		Effect size and 95% CI for indirect treatment	
	Norm-referenced	Criterion-referenced	Norm-referenced	Criterion-referenced
RCTs				
Almost and Rosenbaum, 1997 (n = 26)	+1.37** (+0.52/+2.23)	+1.56** (+0.68/+2.44)		
Lancaster, 1991 (n = 10)		+1.04 (–0.38/+2.36)		–0.01 (–1.25/+1.23)
Methany and Panagos, 1978 (n = 16)	Ti +0.75 (–0.26/+1.76) Tj +0.45 (–0.54/+1.44)			
Shelton *et al,* 1978 (n = 30) (Study 1)			Ti +0.27 (–0.4/+1.00) Tj –0.18 (–0.90/+0.54)	Ti +0.08 (–0.63/+0.80) Tj –0.17 (–0.89/+0.54)
Quasi-experimental				
Fey *et al,* 1994 (n = 18)		+0.20 (–0.73/+1.13)		–0.06 (–1.04/+0.92)
Shelton *et al,* 1978 (Study 2)			Ti –0.08 (–1.39/+1.24) Tj –0.95 (–2.34/+0.44)	Ti –0.31 (–1.63/+1.02) Tj –0.73 (–2.01/+0.55)

** *Significant at* p < *0.01*
Ti and Tj denote that two different treatments were used in the study

TABLE 46 *Summary of weighted ANOVA for articulation/phonology outcome measures (norm-referenced tests) across service delivery*

Between-group variable	*k*	*d*	95% CI	*r*	*p*	*Qw*	*p*
Direct	2	+1.11	+0.46/+1.77	+0.49	0.001	0.85	0.36
Indirect	2	+0.20	–0.44/+0.83	+0.10	0.55	0.21	0.64
Overall	4	+0.64	+0.18/+1.09	+0.30			
			Qb(1) = 3.92, *p* < 0.05				

k *is number of studies in each group*
d *is the effect size*
CI indicate the 95% confidence interval for each effect size
r *is a correlation coefficient and p is its significance and the significance also of the effect size*
Qw *is the test of within-group variation*
Qb *is the test of between-group variation: the figure in brackets represents the degrees of freedom for the comparison*

TABLE 47 *Summary of weighted ANOVA for articulation/phonology outcome measures (criterion-referenced tests) across service delivery*

Between-group variable	*k*	*d*	95% CI	*r*	*p*	*Qw*	*p*
Direct	3	+0.94	+0.37/+1.52	+0.43	0.001	4.36	0.11
Indirect	4	–0.02	–0.51/+0.47	–0.01	0.93	0.26	0.97
Overall	7	+0.38	+0.01/+0.759	+0.19			
			Qb(1) = 6.29, *p* < 0.02				

k *is number of studies in each group*
d *is the effect size*
CI indicate the 95% confidence interval for each effect size
r *is a correlation coefficient and p is its significance and the significance also of the effect size*
Qw *is the test of within-group variation*
Qb *is the test of between-group variation: the figure in brackets represents the degrees of freedom for the comparison*

TABLE 48 *Summary of effect sizes for expressive language outcome measures*

Study	Effect size and 95% CI for direct treatment		Effect size and 95% CI for indirect treatment	
	Norm-referenced	Criterion-referenced	Norm-referenced	Criterion-referenced
RCTs				
Fey *et al,* 1993 (n = 20)	+0.83 (–0.09/+1.74)		+0.89 (–0.06/+1.83)	
Gibbard, 1994 (Study 1) (n = 36)			+2.33** (+1.49/+3.18)	+1.76*** (+0.99/+2.53)
Girolametto *et al,* 1996 (n = 25)			+0.84* (+0.02/+1.66)	+1.00* (+0.17/+1.83)
Girolametto *et al,* 1995 (n = 16)			+0.23 (–0.75/+1.22)	+0.87 (–0.16/+1.89)
Methany and Panagos, 1978 (n = 16)	+1.00 (–0.04/+2.03)			
Schwartz *et al,* 1985 (n = 10)		+0.71 (–0.87/+2.29)		
Quasi-experimental				
Fey *et al,* 1994 (n = 18)	+0.77 (–0.19/+1.74)		+0.43 (–0.56/+1.43)	
Gibbard, 1994 (Study 2) (n = 16)	+0.72 (–0.29/+1.74)	+0.83 (–0.18/+1.854)	+1.33* (+0.29/+2.38)	+1.42* (+0.351/+2.48)
McDade and McCartan, 1996 (n = 18)				+0.47 (–0.46/+1.41)
Stevenson *et al,* 1982 (n = 22)	+0.12 (–0.72/+0.96)			
Ward, 1994 (Group 1) (n = 77)			+1.53** (+1.02/+2.03)	
Ward, 1994 (Group 2) (n = 10)			+1.03 (–0.40/+2.45)	
Whitehurst *et al,* 1991 (n = 62)			+0.66* (+0.136/+1.18)	
Wilcox and Leonard, 1978 (n = 24)		+2.37*** (+1.33/+3.42)		
Zwitman *et al,* 1979 (n = 22)		+0.55 (–0.30/+1.41)		

* *Significant at* p < *0.05;* ** *Significant at* p < *0.01;* *** *Significant at* p < *0.001*

TABLE 49 Summary of weighted ANOVA for expressive language outcome measures (norm-referenced tests) across study design

Between-group variable	***k***	***d***	**95% CI**	***r***	***p***	***Qw***	***p***
RCT	6	+1.07	+0.70/+1.45	+0.47	0.05	12.04	0.06
Quasi-experimental	8	+0.91	+0.64/+1.18	+0.41	0.001	11.72	0.16
Overall	14	+0.97	+0.75/+1.19	+0.44	0.001	24.23	0.0001
			Qb(1) = 0.47, *p* = 0.49, ns				

k *is number of studies in each group*
d *is the effect size*
CI indicate the 95% confidence interval for each effect size
r *is a correlation coefficient and p is its significance and the significance also of the effect size*
Qw *is the test of within-group variation*
Qb *is the test of between-group variation: the figure in brackets represents the degrees of freedom for the comparison*
ns = not significant

TABLE 50 *Summary of weighted ANOVA for expressive language outcome measures (norm-referenced tests) across service delivery*

Between-group variable	***k***	***d***	**95% CI**	***r***	***p***	***Qw***	***p***
Direct	5	+0.65	+0.23/+1.10	+0.31	0.05	2.22	0.82
Indirect	9	+1.08	+0.83/+1.34	+0.48	0.001	19.09	0.02
Overall	14	+0.97	+0.75/+1.19	+0.44	0.0001	24.23	0.0001
			Qb(1) = 2.92, *p* = 0.09, ns				

k *is number of studies in each group*
d *is the effect size*
CI indicate the 95% confidence interval for each effect size
r *is a correlation coefficient and p is its significance and the significance also of the effect size*
Qw *is the test of within-group variation*
Qb *is the test of between-group variation: the figure in brackets represents the degrees of freedom for the comparison*
ns = not significant

TABLE 51 *Summary of effect sizes for receptive language outcome measures*

Study	Effect size and 95% CI for direct treatment		Effect size and 95% CI for indirect treatment	
	Norm-referenced	Criterion-referenced	Norm-referenced	Criterion-referenced
RCTs				
Shelton *et al,* 1978 (n = 30)			Ti +0.16 (–0.56/+0.88) Tj +0.30 (–0.42/+1.02)	
Gibbard, 1994 (Study 1) (n = 36)			+1.49*** (+0.75/+2.23)	
Quasi-experimental				
Stevenson *et al,* 1982 (n = 21)	+0.12 (–0.74/+0.98)			
Ward, 1994 (Group 1) (n = 84)			+2.19*** (+1.65/+2.74)	
Ward, 1994 (Group 2) (n = 13)			+1.92** (+0.53/+3.31)	
Gibbard, 1994 (Study 2) (n = 16)	–0.16 (–0.88/+0.56)		+0.68 (–0.30/+1.66)	

* *Significant at* p < *0.05;* ** *Significant at* p < *0.01;* *** *Significant at* p < *0.001*

TABLE 52 *Summary of weighted ANOVA for receptive language outcome measures (norm-referenced tests) across service delivery*

Between-group variable	*k*	*d*	95% CI	*r*	*p*	*Qw*	*p*
Direct	2	–0.02	–0.66/+0.63	–0.009	0.05	0.23	0.89
Indirect	5	+1.43	+1.09/+1.77	+0.58	0.001	20.00	0.001
Overall	7	+1.12	+0.82/+1.42	+0.49	0.0001	24.23	0.0001
Qb(1) = 15.20, *p* = 0.0001							

k *is number of studies in each group*
d *is the effect size*
CI indicate the 95% confidence interval for each effect size
r *is a correlation coefficient and* p *is its significance and the significance also of the effect size*
Qw *is the test of within-group variation*
Qb *is the test of between-group variation: the figure in brackets represents the degrees of freedom for the comparison*

HTA panel membership

This report was identified as a priority by the Population Screening Panel.

Acute Sector Panel

Chair: Professor John Farndon, University of Bristol †

Professor Senga Bond, University of Newcastle-upon-Tyne †

Professor Ian Cameron, Southeast Thames Regional Health Authority

Ms Lynne Clemence, Mid-Kent Health Care Trust †

Professor Francis Creed, University of Manchester †

Professor Cam Donaldson, University of Aberdeen

Mr John Dunning, Papworth Hospital, Cambridge †

Professor Richard Ellis, St James's University Hospital, Leeds

Mr Leonard Fenwick, Freeman Group of Hospitals, Newcastle-upon-Tyne †

Dr David Field, Leicester Royal Infirmary †

Ms Grace Gibbs, West Middlesex University Hospital NHS Trust †

Dr Neville Goodman, Southmead Hospital Services Trust, Bristol †

Mr Ian Hammond, Bedford & Shires Health & Care NHS Trust

Professor Adrian Harris, Churchill Hospital, Oxford

Professor Robert Hawkins, University of Bristol †

Dr Chris McCall, General Practitioner, Dorset †

Professor Alan McGregor, St Thomas's Hospital, London

Mrs Wilma MacPherson, St Thomas's & Guy's Hospitals, London

Professor Jon Nicoll, University of Sheffield †

Professor John Norman, University of Southampton

Dr John Pounsford, Frenchay Hospital, Bristol †

Professor Gordon Stirrat, St Michael's Hospital, Bristol

Professor Michael Sheppard, Queen Elizabeth Hospital, Birmingham †

Dr William Tarnow-Mordi, University of Dundee

Professor Kenneth Taylor, Hammersmith Hospital, London

Diagnostics and Imaging Panel

Chair: Professor Mike Smith, University of Leeds †

Professor Michael Maisey, Guy's & St Thomas's Hospitals, London *

Professor Andrew Adam, UMDS, London †

Dr Pat Cooke, RDRD, Trent Regional Health Authority

Ms Julia Davison, St Bartholomew's Hospital, London †

Professor Adrian Dixon, University of Cambridge †

Professor MA Ferguson-Smith, University of Cambridge †

Dr Mansel Hacney, University of Manchester

Professor Sean Hilton, St George's Hospital Medical School, London

Mr John Hutton, MEDTAP International Inc., London †

Professor Donald Jeffries, St Bartholomew's Hospital, London †

Dr Andrew Moore, Editor, *Bandolier* †

Professor Chris Price, London Hospital Medical School †

Dr Ian Reynolds, Nottingham Health Authority

Professor Colin Roberts, University of Wales College of Medicine †

Miss Annette Sergeant, Chase Farm Hospital, Enfield

Professor John Stuart, University of Birmingham

Dr Ala Szczepura, University of Warwick †

Mr Stephen Thornton, Cambridge & Huntingdon Health Commission

Dr Gillian Vivian, Royal Cornwall Hospitals Trust †

Dr Jo Walsworth-Bell, South Staffordshire Health Authority †

Dr Greg Warner, General Practitioner, Hampshire †

Methodology Panel

Chair: Professor Martin Buxton, Brunel University †

Professor Anthony Culyer, University of York *

Dr Doug Altman, Institute of Health Sciences, Oxford †

Professor Michael Baum, Royal Marsden Hospital

Professor Nick Black, London School of Hygiene & Tropical Medicine †

Professor Ann Bowling, University College London Medical School †

Dr Rory Collins, University of Oxford

Professor George Davey-Smith, University of Bristol

Dr Vikki Entwistle, University of Aberdeen †

Professor Ray Fitzpatrick, University of Oxford †

Professor Stephen Frankel, University of Bristol

Dr Stephen Harrison, University of Leeds

Mr Philip Hewitson, Leeds FHSA

Professor Richard Lilford, Regional Director, R&D, West Midlands †

Mr Nick Mays, King's Fund, London †

Professor Ian Russell, University of York †

Professor David Sackett, Centre for Evidence Based Medicine, Oxford †

Dr Maurice Slevin, St Bartholomew's Hospital, London

Dr David Spiegelhalter, Institute of Public Health, Cambridge †

Professor Charles Warlow, Western General Hospital, Edinburgh †

* Previous Chair
† Current members

continued

continued

Pharmaceutical Panel

Chair: Professor Tom Walley, University of Liverpool †

Professor Michael Rawlins, University of Newcastle-upon-Tyne*

Dr Colin Bradley, University of Birmingham

Professor Alasdair Breckenridge, RDRD, Northwest Regional Health Authority

Ms Christine Clark, Hope Hospital, Salford †

Mrs Julie Dent, Ealing, Hammersmith & Hounslow Health Authority, London

Mr Barrie Dowdeswell, Royal Victoria Infirmary, Newcastle-upon-Tyne

Dr Desmond Fitzgerald, Mere, Bucklow Hill, Cheshire

Dr Alistair Gray, Health Economics Research Unit, University of Oxford †

Professor Keith Gull, University of Manchester

Dr Keith Jones, Medicines Control Agency

Professor Trevor Jones, ABPI, London †

Ms Sally Knight, Lister Hospital, Stevenage †

Dr Andrew Mortimore, Southampton & SW Hants Health Authority †

Mr Nigel Offen, Essex Rivers Healthcare, Colchester †

Dr John Posnett, University of York

Mrs Marianne Rigge, The College of Health, London †

Mr Simon Robbins, Camden & Islington Health Authority, London †

Dr Frances Rotblat, Medicines Control Agency †

Mrs Katrina Simister, Liverpool Health Authority †

Dr Ross Taylor, University of Aberdeen †

Dr Tim van Zwanenberg, Northern Regional Health Authority

Dr Kent Woods, RDRD, Trent RO, Sheffield †

Population Screening Panel

Chair: Professor Sir John Grimley Evans, Radcliffe Infirmary, Oxford †

Dr Sheila Adam, Department of Health*

Ms Stella Burnside, Altnagelvin Hospitals Trust, Londonderry †

Dr Carol Dezateux, Institute of Child Health, London †

Dr Anne Dixon Brown, NHS Executive, Anglia & Oxford †

Professor Dian Donnai, St Mary's Hospital, Manchester †

Dr Tom Fahey, University of Bristol †

Mrs Gillian Fletcher, National Childbirth Trust †

Professor George Freeman, Charing Cross & Westminster Medical School, London

Dr Mike Gill, Brent & Harrow Health Authority †

Dr JA Muir Gray, RDRD, Anglia & Oxford RO †

Dr Ann Ludbrook, University of Aberdeen †

Professor Alexander Markham, St James's University Hospital, Leeds †

Professor Theresa Marteau, UMDS, London

Dr Ann McPherson, General Practitioner, Oxford †

Professor Catherine Peckham, Institute of Child Health, London

Dr Connie Smith, Parkside NHS Trust, London

Dr Sarah Stewart-Brown, University of Oxford †

Ms Polly Toynbee, Journalist †

Professor Nick Wald, University of London †

Professor Ciaran Woodman, Centre for Cancer Epidemiology, Manchester

Primary and Community Care Panel

Chair: Dr John Tripp, Royal Devon & Exeter Healthcare NHS Trust †

Professor Angela Coulter, King's Fund, London *

Professor Martin Roland, University of Manchester *

Dr Simon Allison, University of Nottingham

Mr Kevin Barton, East London & City Health Authority †

Professor John Bond, University of Newcastle-upon-Tyne †

Ms Judith Brodie, Age Concern, London †

Dr Nicky Cullum, University of York †

Professor Shah Ebrahim, Royal Free Hospital, London

Mr Andrew Farmer, Institute of Health Sciences, Oxford †

Ms Cathy Gritzner, The Patients' Association †

Professor Andrew Haines, RDRD, North Thames Regional Health Authority

Dr Nicholas Hicks, Oxfordshire Health Authority †

Professor Richard Hobbs, University of Birmingham †

Professor Allen Hutchinson, University of Sheffield †

Mr Edward Jones, Rochdale FHSA

Professor Roger Jones, UMDS, London

Mr Lionel Joyce, Chief Executive, Newcastle City Health NHS Trust

Professor Martin Knapp, London School of Economics & Political Science

Professor Karen Luker, University of Liverpool

Professor David Mant, NHS Executive South & West †

Dr Fiona Moss, North Thames British Postgraduate Medical Federation †

Professor Dianne Newham, King's College London

Professor Gillian Parker, University of Leicester †

Dr Robert Peveler, University of Southampton †

Dr Mary Renfrew, University of Oxford

Ms Hilary Scott, Tower Hamlets Healthcare NHS Trust, London †

* Previous Chair
† Current members

Health Technology Assessment 1998; Vol. 2: No. 9

Copies of this report can be obtained from:

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 1703 595 639 Email: hta@soton.ac.uk
http://www.soton.ac.uk/~hta

ISSN 1366-5278